



US006128624A

**United States Patent** [19][11] **Patent Number:** **6,128,624****Papierniak et al.**[45] **Date of Patent:** **\*Oct. 3, 2000**

[54] **COLLECTION AND INTEGRATION OF INTERNET AND ELECTRONIC COMMERCE DATA IN A DATABASE DURING WEB BROWSING**

5,812,668 9/1998 Weber ..... 380/24  
 5,819,092 10/1998 Ferguson et al. .... 395/701  
 5,862,223 1/1999 Walker et al. .... 380/25

**OTHER PUBLICATIONS**

[75] **Inventors:** Karen A. Papierniak, St. Paul, Minn.; James E. Thaisz, Lincroft; Luo-Jen Chlang, Freehold, both of N.J.

Akashi, O, et al., Information distribution by fleamarket system, IEEE 1996, 139-146, Jun. 1996.

Rosenthal, D, EINet: a secure, open network for electronic commerce, IEEE 1994, 219-226, Dec. 1994.

[73] **Assignee:** NCR Corporation, Dayton, Ohio

*Primary Examiner*—Jack M. Choules

*Assistant Examiner*—Srirama Channavajjala

*Attorney, Agent, or Firm*—Lowe Hauptman Gopstein Gilman & Berner

[\*] **Notice:** This patent issued on a continued prosecution application filed under 37 CFR 1.53(d), and is subject to the twenty year patent term provisions of 35 U.S.C. 154(a)(2).

[57] **ABSTRACT**

A system integrates data elements from an operational database of an Internet Service Provider (ISP) and Commerce Service Provider (CSP) into a predetermined format for supporting collection of the Internet and/or electronic commerce data. The system includes a database storing the electronic commerce data for interrogation by the ISP/CSP, a user station, and a design database. The design database includes a logical data model providing a description of the database facilitating the integration of the Internet and/or electronic commerce data and facilitating query and report access of the database. The design database also includes a formatted file library providing classifications including at least one of process characterizations, customer descriptions, preference determinations, and behavior patterns, the classifications being reusable for at least one of different technical processes and different customer problems. An optional user interface is also provided which is capable of providing the functions of query access and report access of the design database.

[21] **Appl. No.:** 08/969,082

[22] **Filed:** Nov. 12, 1997

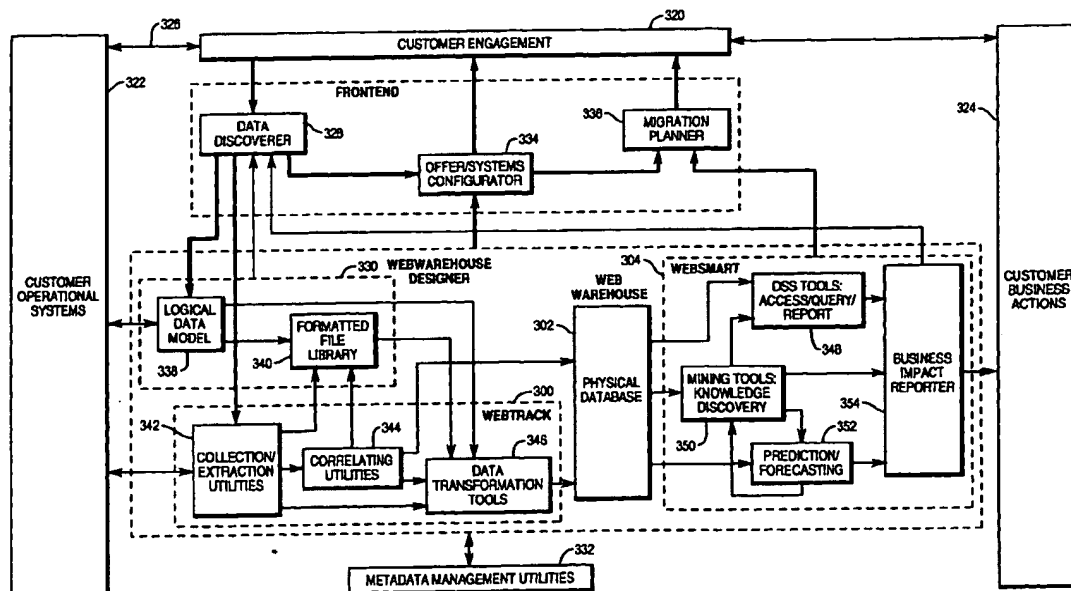
[51] **Int. Cl.<sup>7</sup>** ..... G06F 17/30

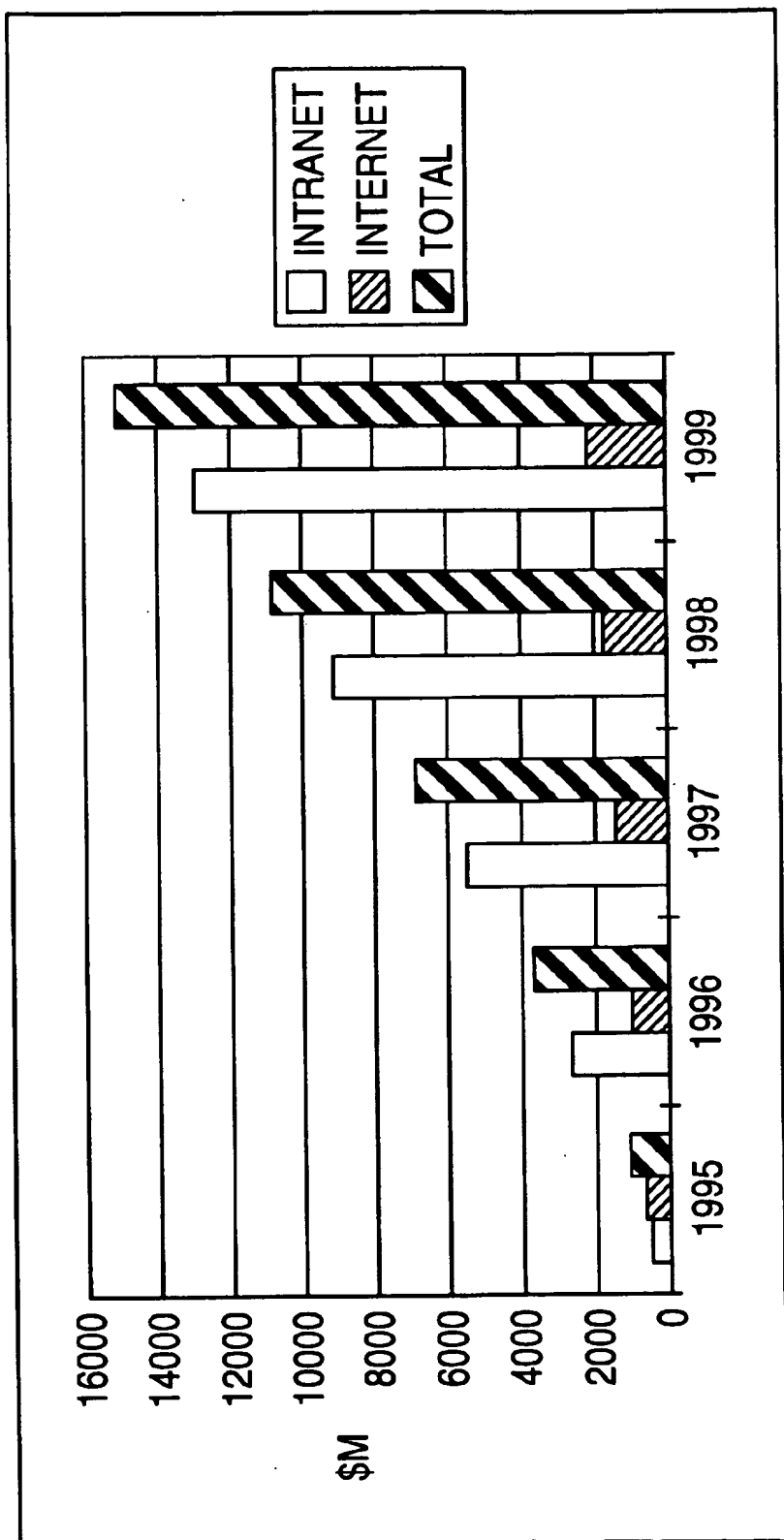
[52] **U.S. Cl.** ..... 707/104; 707/10; 707/102; 705/10; 705/26; 705/27

[58] **Field of Search** ..... 707/1-5, 100-104, 707/10, 200-204, 513, 501, 503; 705/35, 37, 39, 16, 26, 27, 21, 10; 706/45, 46, 47; 395/200.33, 200.48, 200.34; 380/9, 21, 22, 23, 24, 25, 201; 709/217-219

[56] **References Cited****U.S. PATENT DOCUMENTS**

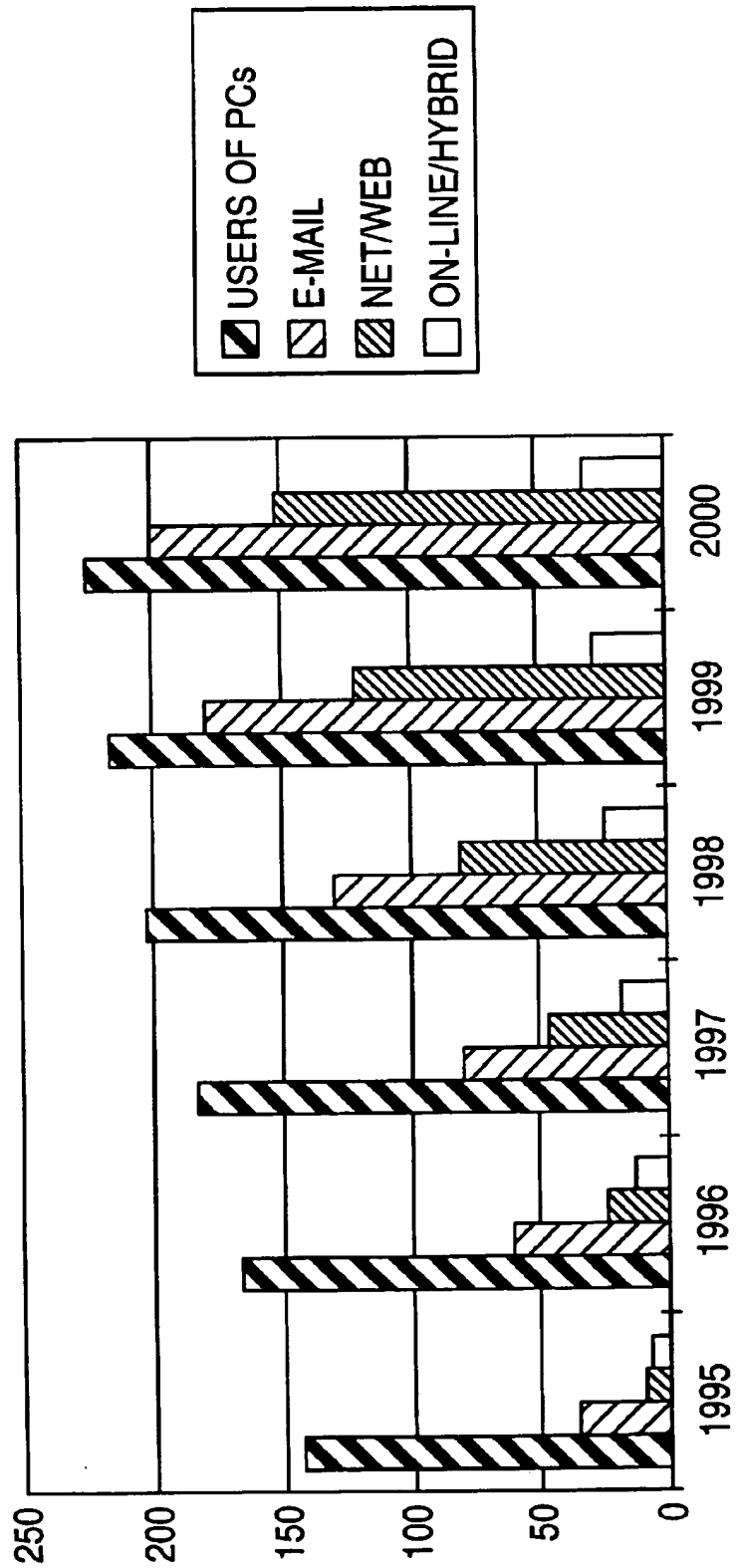
5,671,279 9/1997 Elgamal ..... 380/23  
 5,701,400 12/1997 Amado ..... 706/45  
 5,710,887 1/1998 Chelliah et al. .... 705/26  
 5,729,594 3/1998 Klingman ..... 379/93.12  
 5,809,144 9/1998 Sirbu et al. .... 380/25

**29 Claims, 22 Drawing Sheets**

**FIG. 1****WORLD WIDE WEB SERVER SALES FORECAST**

**FIG. 2**

WEB USERS  
(ESTIMATED GROWTH IN MILLIONS)



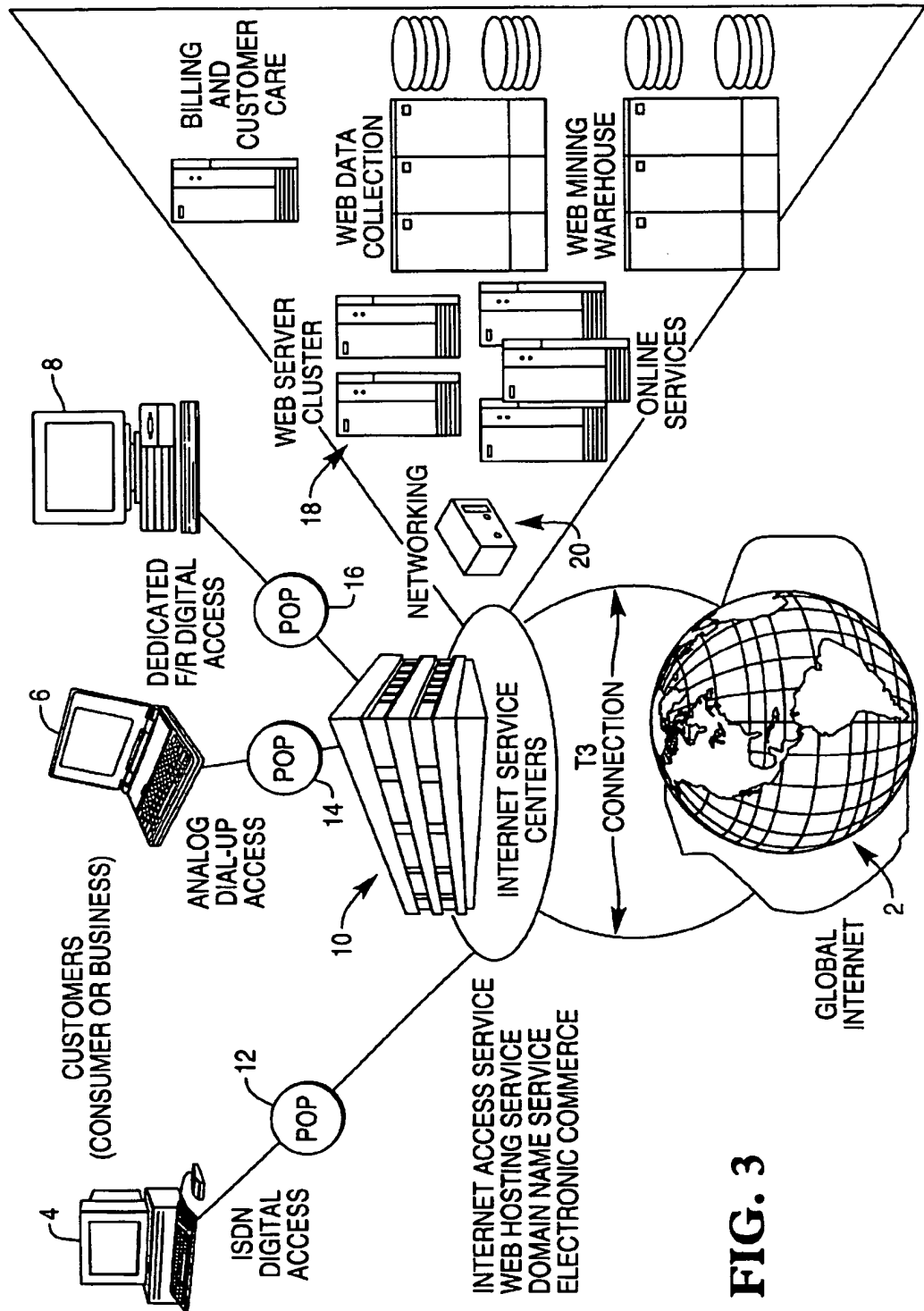
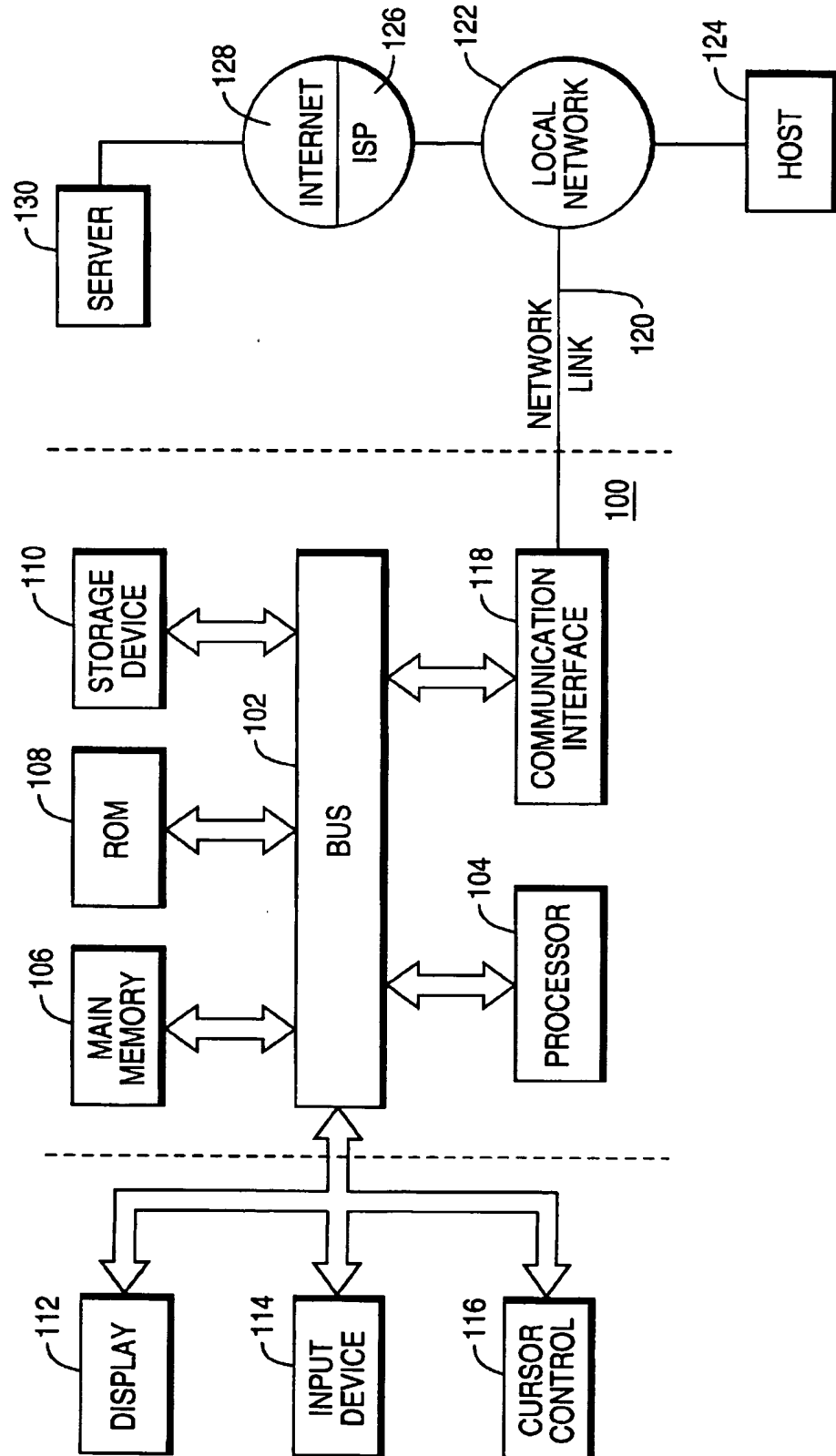
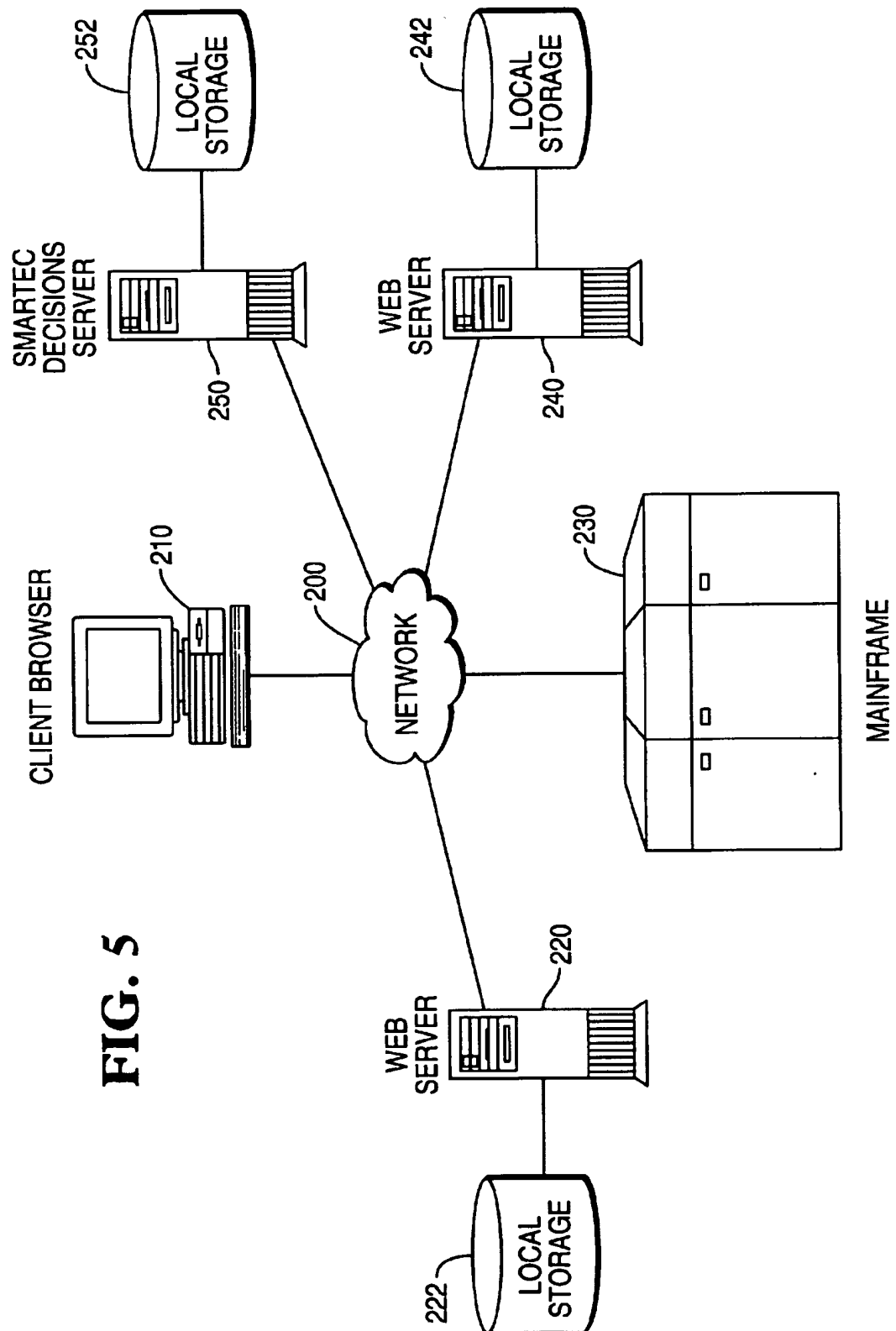
**FIG. 3**

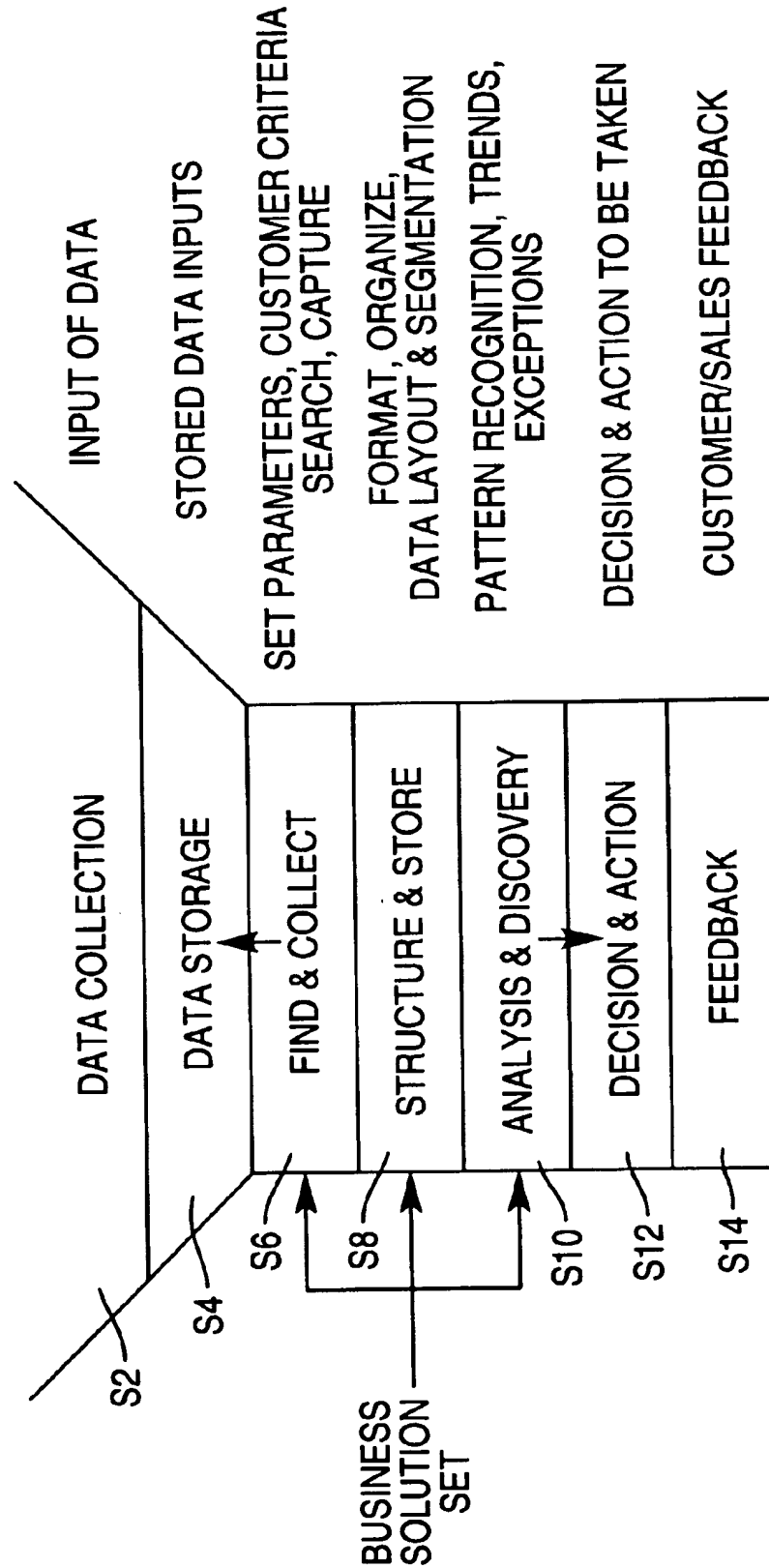
FIG. 4

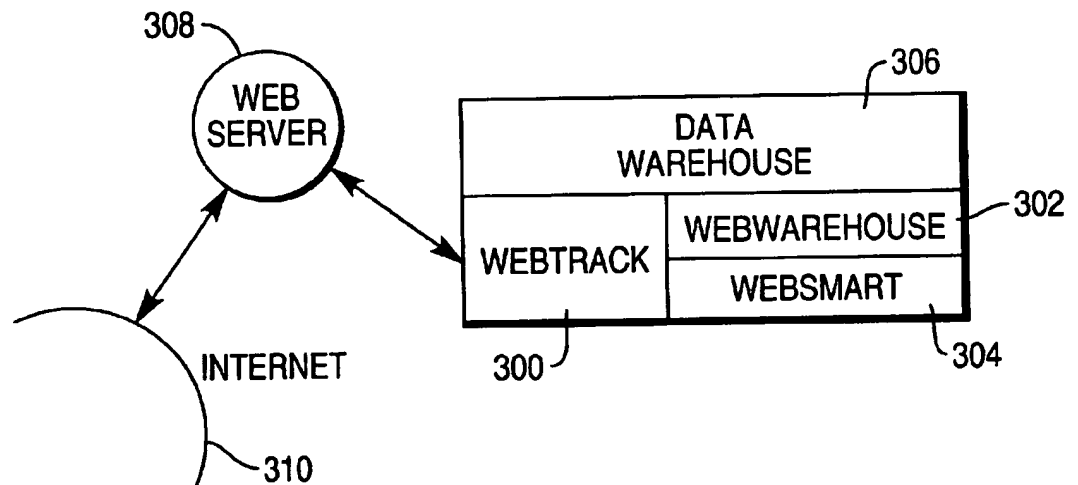




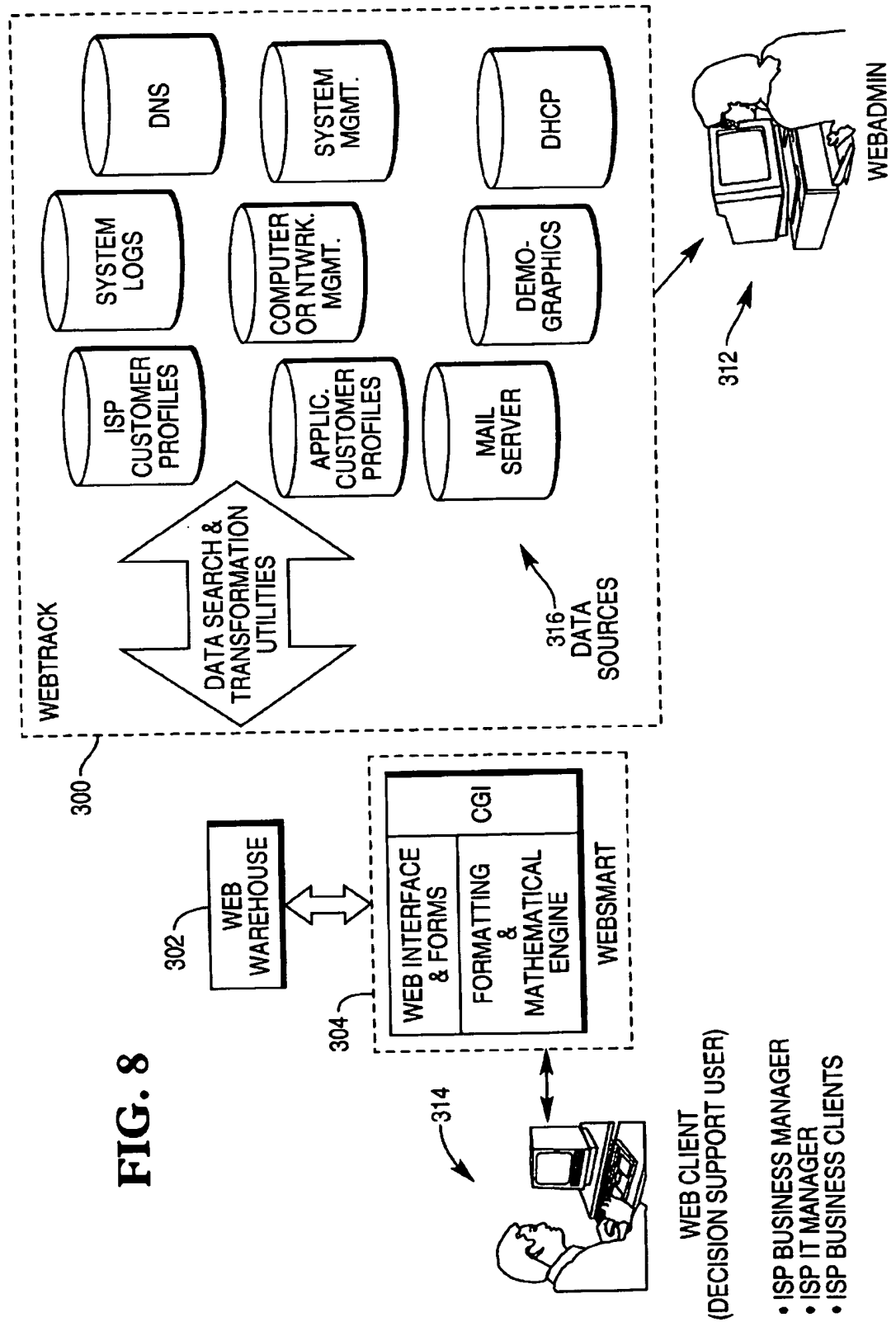
**FIG. 5**

**FIG. 6**



**FIG. 7**





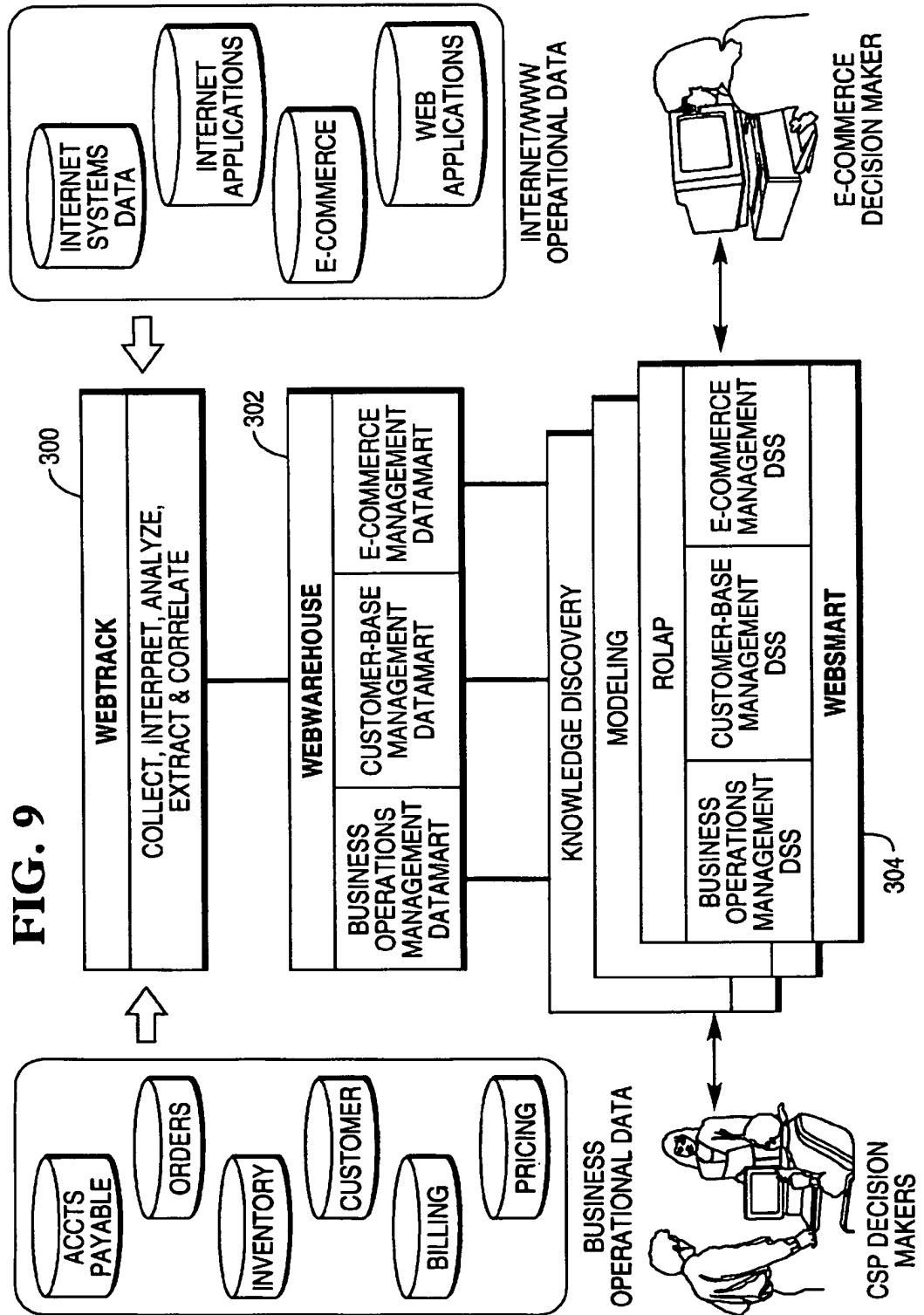
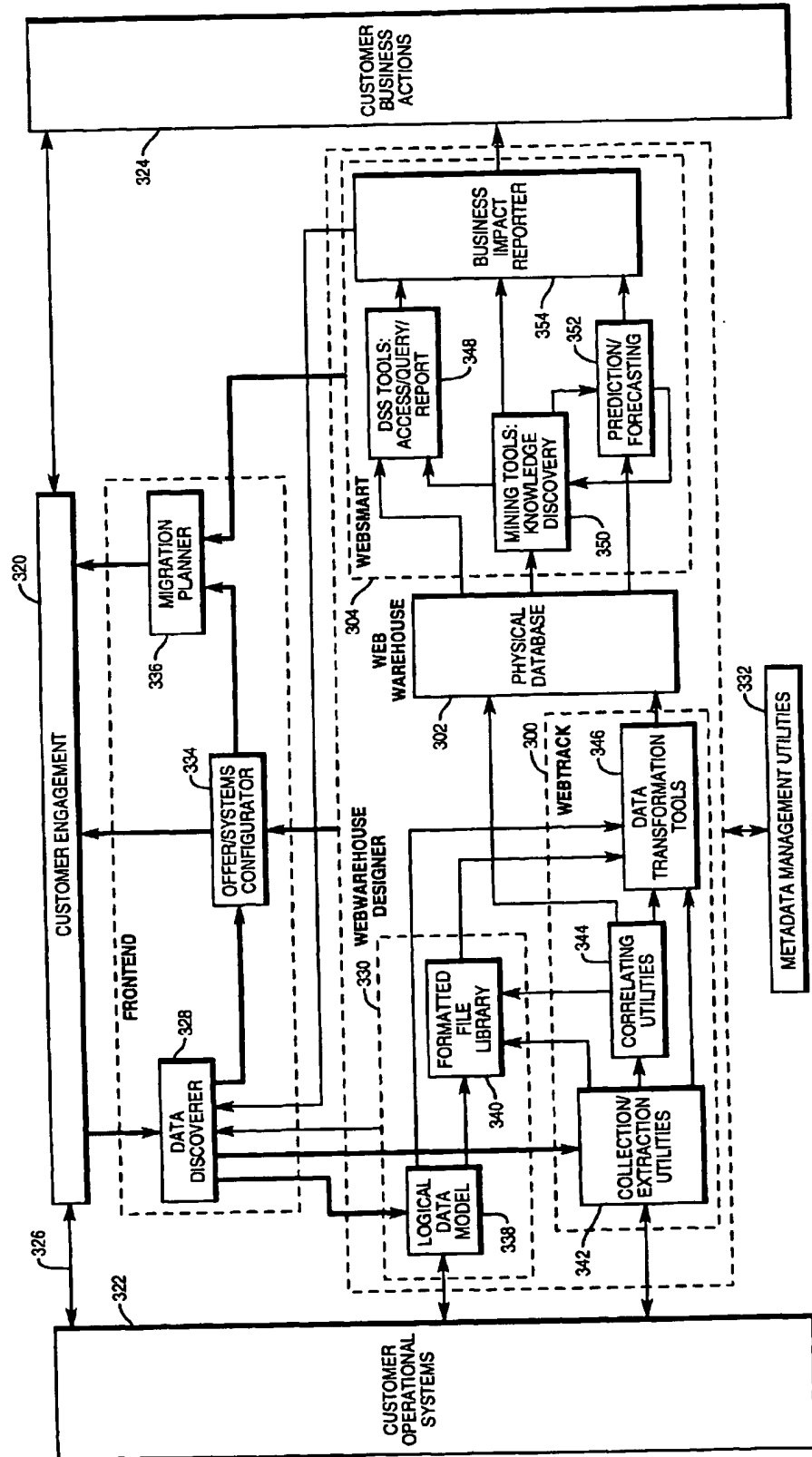
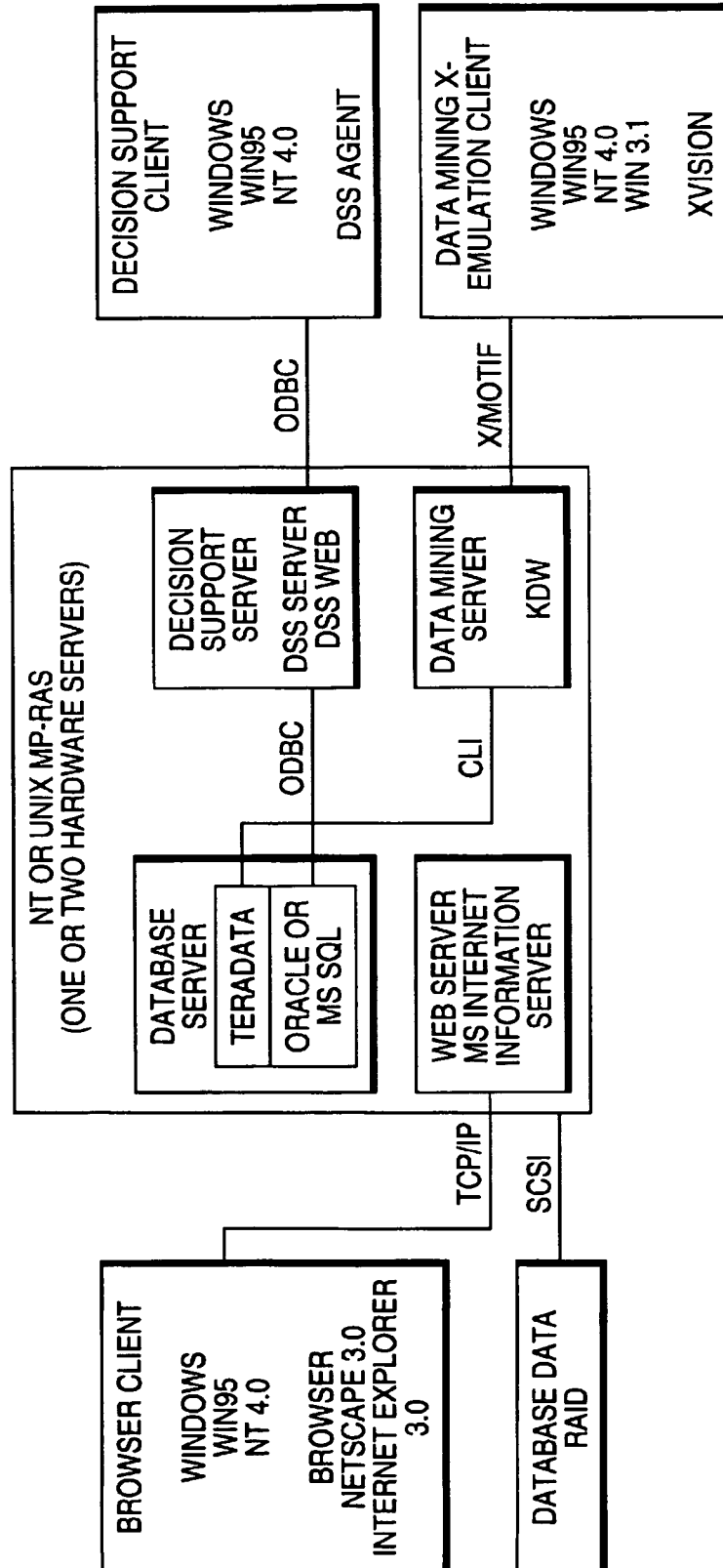


FIG. 10

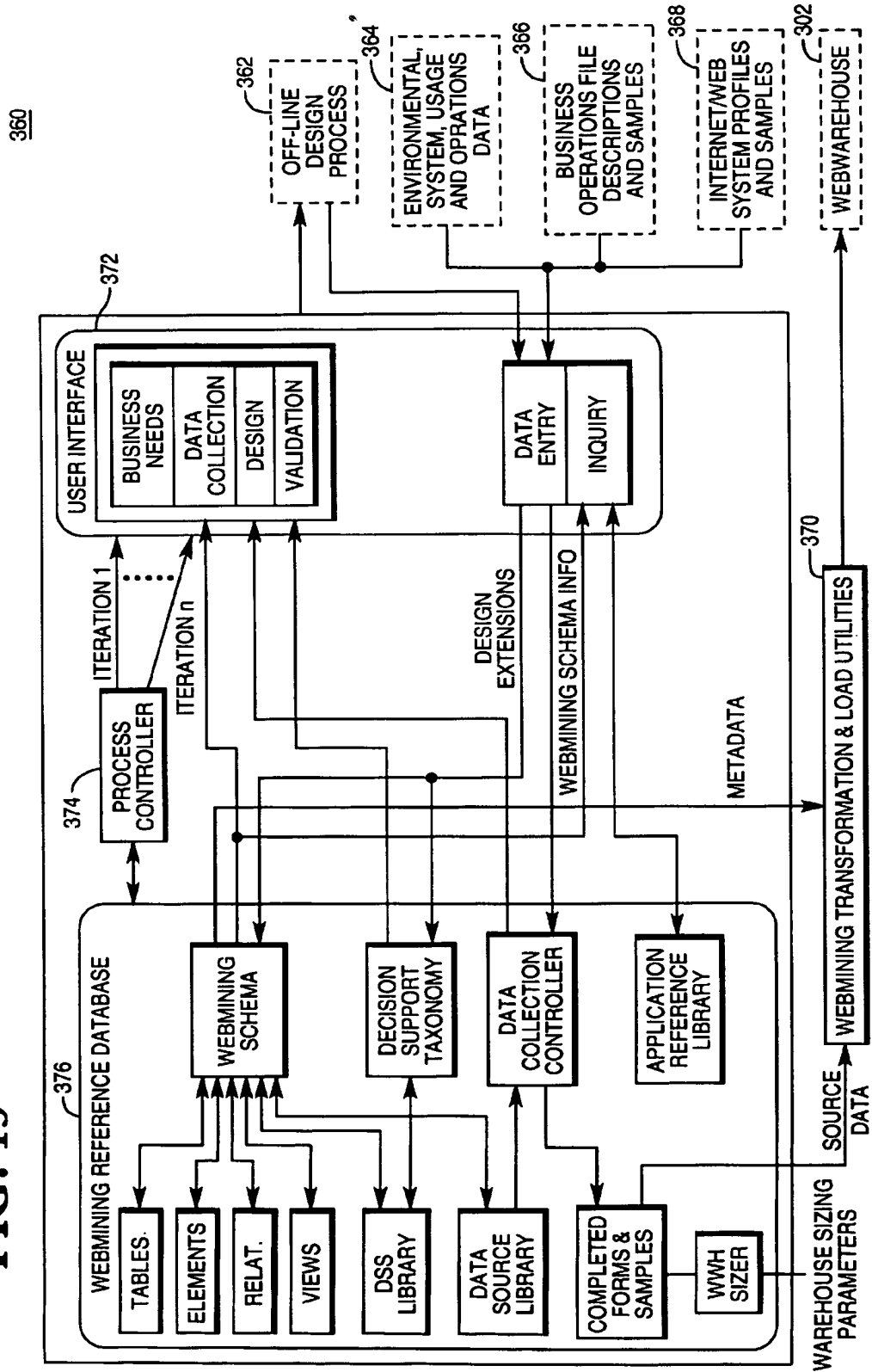


**FIG. 11**

**FIG. 12**

<b>Hardware Options</b> (minimum)	<p><b>WM UNIX Platform</b> - 2 Pentiums 166MHz CPUs, 128MB of base memory plus 32MB per user up to 5 users.</p> <p><b>WM NT Server Platform</b> - 2 Pentiums 166MHz CPUs, 128MB base memory plus 32MB per user up to 5 users.</p> <p><b>Browser, DSS, Data Mining Client Platform</b> - Any PC, Laptop, Server connected to the WM UNIX or NT server. Data Mining client hardware must have 1024x768 256 color graphics capability.</p>
<b>Software Base Options</b> (minimum)	<p><b>UNIX Server</b> - UNIX SVR4 MP-RAS v3.0 or greater.</p> <p><b>NT Server</b> - Microsoft NT v4.0.</p> <p><b>Browser, DSS and Data Mining Client</b> - Windows 95, NT v4.0, Netscape 3.0 or Internet Explorer 3.0, ODBC, and Xvision.</p>
<b>Database Software Options</b> (minimum)	<p>Teradata V2 R1.1.1 or greater for NCR UNIX SVR4 MP-RAS</p> <p>Oracle for Microsoft NT</p> <p>Microsoft SQL Server on Microsoft NT</p>
<b>Database Connectivity Software</b> (minimum)	<p>Teradata CLI R4.0 for NCR UNIX SVR5 MP-RAS</p> <p>Teradata ODBC R2.02 for Microsoft NT</p> <p>ODBC included in Oracle for NT</p> <p>ODBC included in MS SQL Server</p> <p>Both 16 and 32 bit ODBC drivers for Windows 95 and NT 4.0 clients</p>
<b>Information Access or Decision Support and Data Mining Software</b>	<p><b>Info Access/DSS</b> - Platinum Technology, Inc. InfoBeacon, MicroStrategy DSS Agent, Server, Web; Information Advantage Inc. Decision Suite</p> <p><b>Data Mining</b> - Angoss Software International KnowledgeSEEKER, NCR Knowledge Discovery Workbench, SABRE Decision Technologies Datawise</p>
<b>WebMining Tools</b>	<p>Examples of Data Source descriptions, business rules, and reports, Logical Data Models Utilities for Extraction Utilities and Data Transformation Base Structures and schema for each DB supported.</p>

FIG. 13



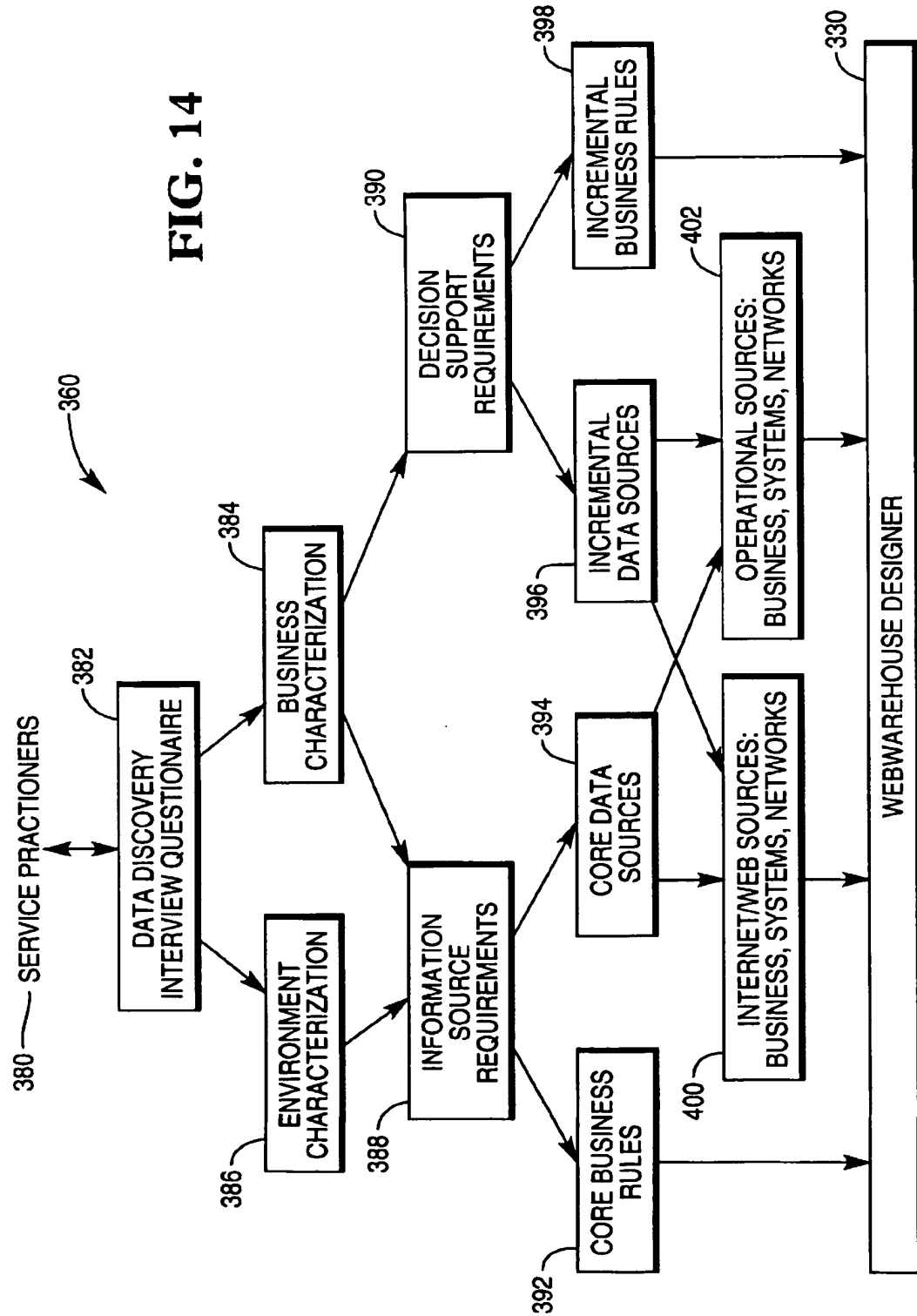
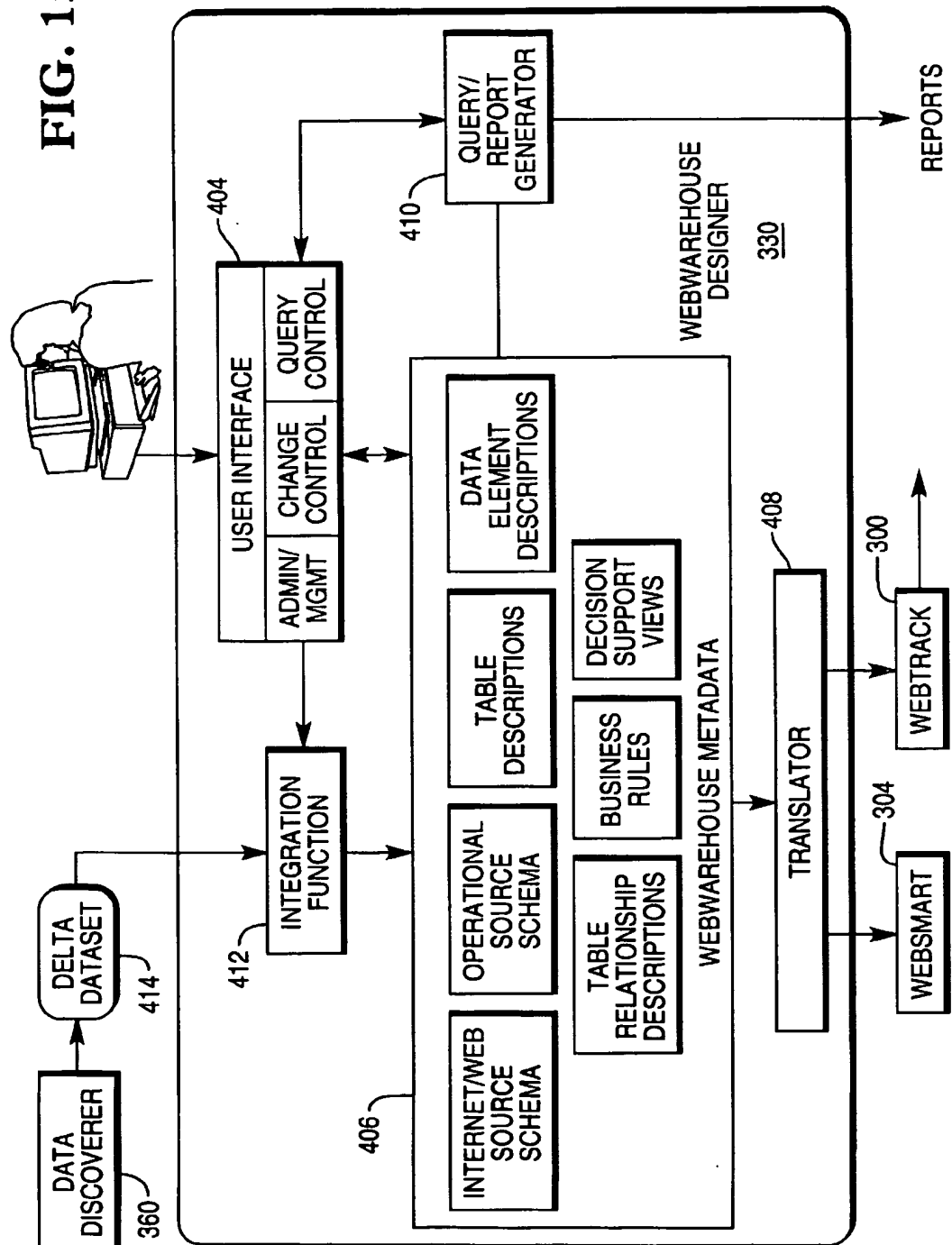
**FIG. 14**

FIG. 15





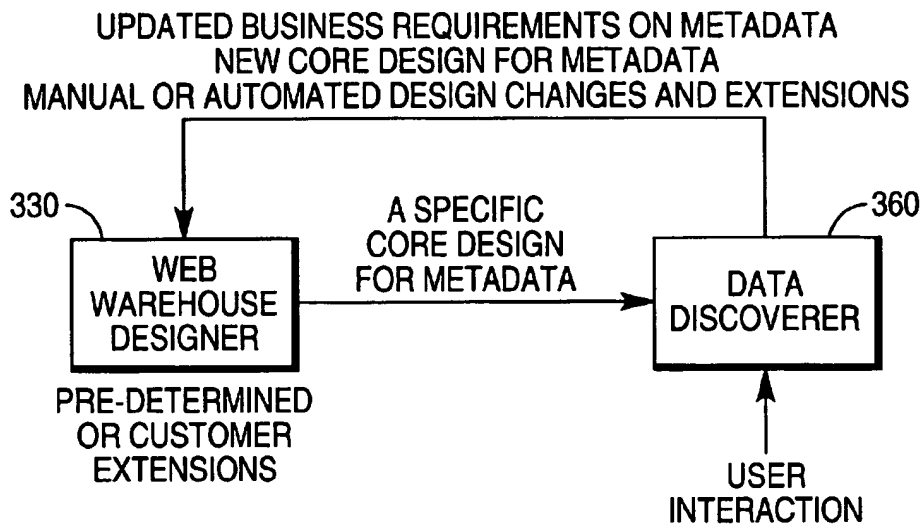
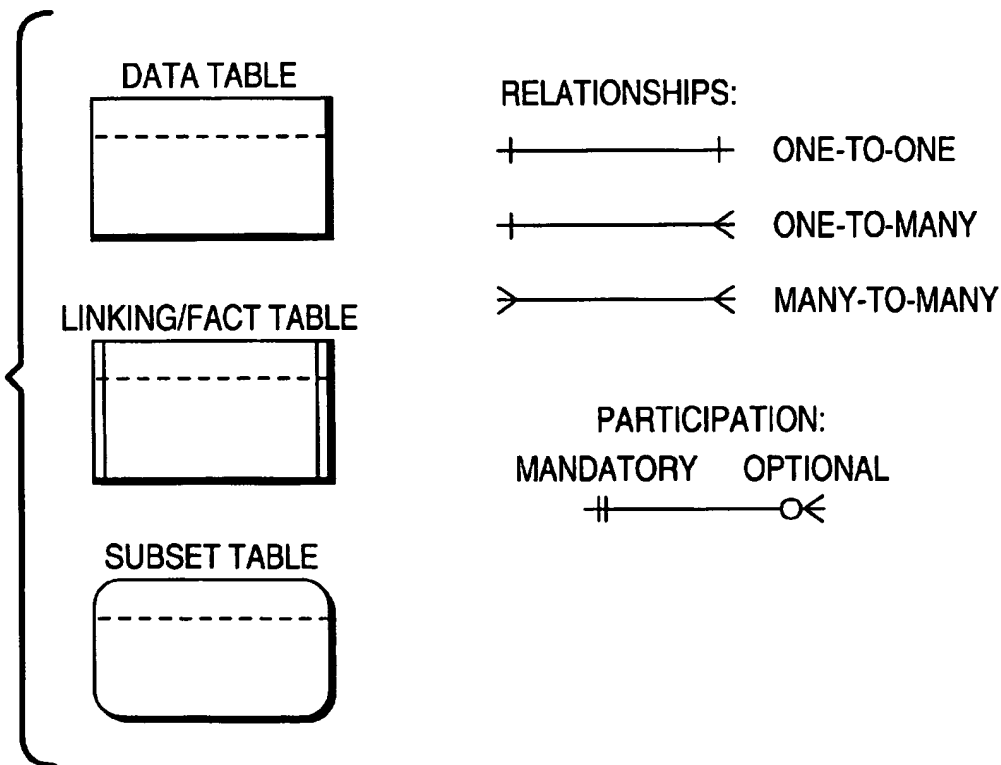
**FIG. 15A****FIG. 17**

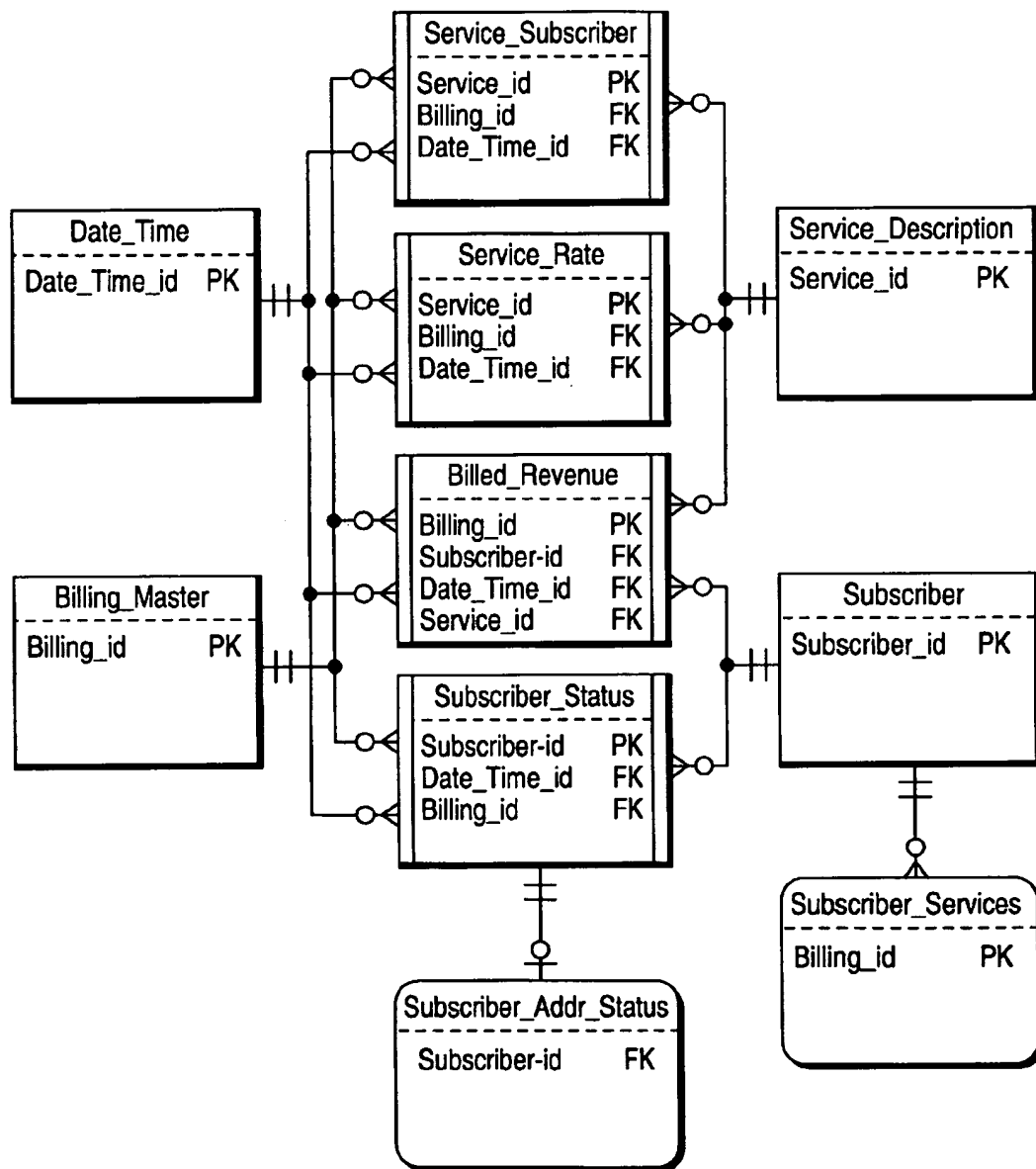
FIG. 16

	CUSTOMER BUSINESS SYSTEMS				INTERNET/WEB SYSTEMS				REFERENCE SYSTEMS			
	BILLING SYSTEM	SUBSCRIBER INFORMATION	BUSINESS CLIENT INFORMATION	ACCESS SERVICES DESCRIPTION	HOSTING/CONSULTING SERVICE DESCRIPTION	INVOICES	POP PROFILES	REMOTE ACCESS SERVER PROFILES	REMOTE ACCESS SERVER LOGS	SERVER ACCESS LOGS	WEB APPLICATION DESCRIPTIONS	SUBSCRIBER BASE DEMOGRAPHICS
												INDIVIDUAL SUBSCRIBER DEMOGRAPHICS
												WEBSITE CONTENT TAXONOMY
												PREFERENCE/INTEREST TAXONOMY
<b>SERVICES PORTFOLIO MANAGEMENT</b>												
<b>ACCESS SERVICES</b>												
SESSION MANAGEMENT							X		X			
REVENUE MANAGEMENT	X	X		X								
SUBSCRIBER MANAGEMENT		X		X								
SERVICES DEMOGRAPHICS		X										X
<b>HOSTING/CONSULTING SERVICES</b>												
REVENUE MANAGEMENT	X		X			X						
BUSINESS CLIENT MANAGEMENT			X									
<b>SUBSCRIBER BASE MANAGEMENT</b>												
<b>SUBSCRIBER VALUE</b>												
SUBSCRIBER PROFITABILITY MANAGEMENT		X				X			X	X		
SUBSCRIBER VALUE-BASED MARKETING		X				X	X		X	X	X	X
<b>SUBSCRIBER PROFILE</b>												
SESSION PROFILE MANAGEMENT		X					X	X	X			
PREFERENCES/INTERESTS-BASED MARKETING		X								X	X	X
DEMOGRAPHICS-BASED MARKETING		X					X					X
<b>APPLICATION MANAGEMENT</b>												
<b>ON-LINE CONTENT EFFECTIVENESS</b>												
CONTENT ACTIVITY MANAGEMENT							X			X	X	X
WEB PROGRAMMING MANAGEMENT										X	X	X

**FIG. 18**

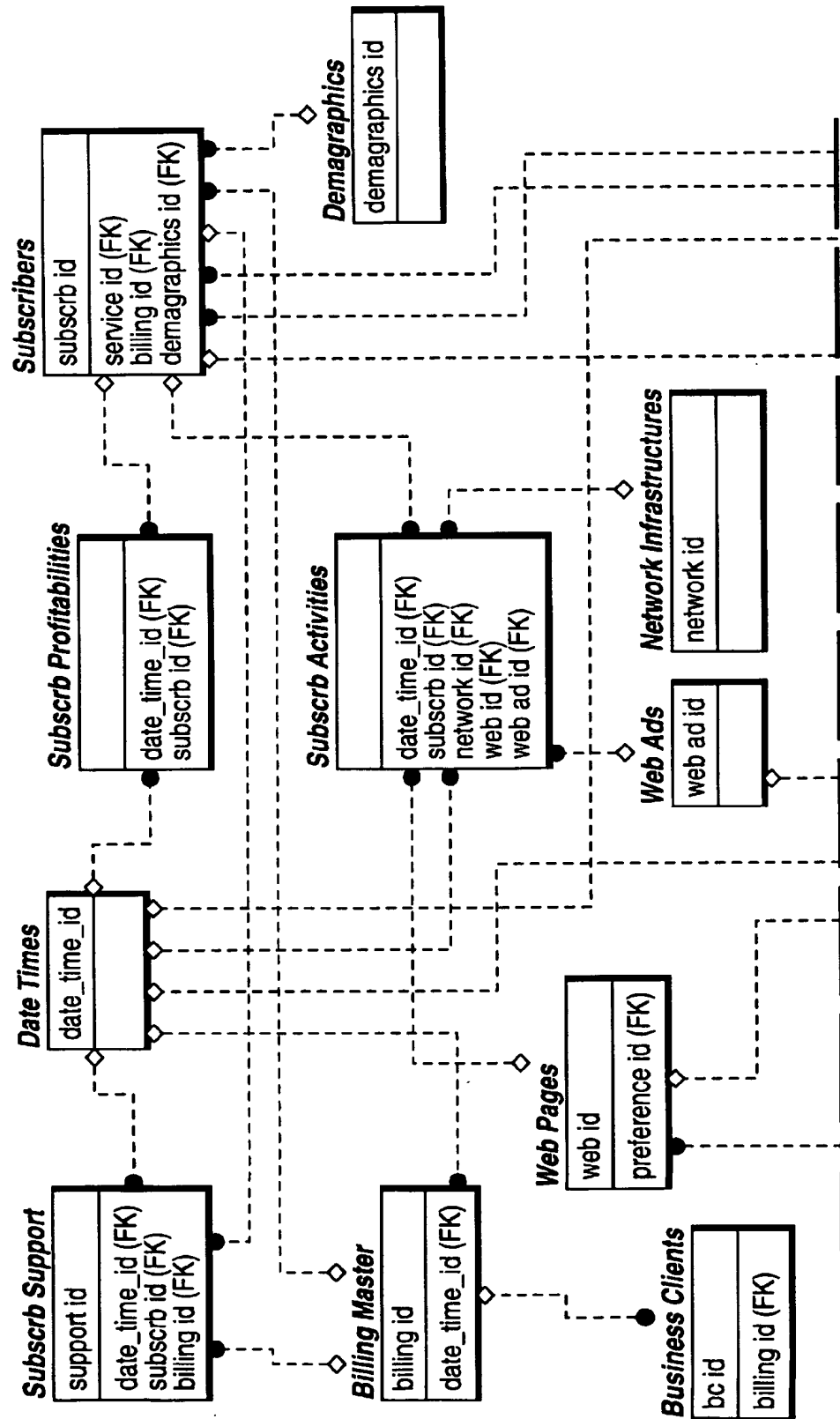
<b>TABLE TYPE</b>	<b>DEFINITION</b>
<b>FACT</b>	A RELATIONSHIP TABLE THAT LINKS TWO OR MORE ENTITIES INTO A MANY-TO-MANY RELATIONSHIP.
<b>DIMENSION</b>	A DATA TABLE THAT DESCRIBES EACH ENTITY.
<b>LOOKUP</b>	AN EXTENSION TO A DIMENSION TABLE USED TO EXPEDITE INDEXING ON ALTERNATE KEYS.
<b>SUBSET</b>	AN EXTENSION TO A DIMENSION TABLE USED TO STORE OPTIONAL DATA THAT MAY NOT PERTAIN TO EVERY OCCURRENCE OF THE RELATED ENTITY.

FIG. 19



**FIG. 20**

<b>Business Operation Management - Subscriber Services Metrics</b>		
<b>Table Name</b>	<b>Table Type</b>	<b>Description</b>
Service_Subscriber	Fact	Facilitates tracking the number of subscribers for each dial-up or dedicated line service over a range of time periods. Enables management to track trends in the customer base as a result of business practices or competitive influences.
Service_Rate	Fact	Facilitates tracking billing rates for each dial-up or dedicated line service over a range of time periods. Enables billed revenue to be calculated based on rates effective at the given time. Important for comparing revenue over various periods of time.
Billed_Revenue	Fact	Tracks individual billing transactions for each subscriber at the time of billing. Used to calculate total billed revenue on a per subscriber basis for any point in time for calculating profitability indices. E.g. Most/least profitable customers.
Subscriber_Status	Fact	Indicates the subscriber's status (e.g. active, terminated) over the time dimension.
Subscriber_Addr_Status	Subset	Used for subscriber address changes over time. The subscriber record contains the current address. This record indicates past addresses.
Date_time	Dimension	Allows queries from various perspectives across the time dimension. Currently, the granularity for this feature set is one month.
Billing_Master	Dimension	Describes the types of transactions for which bills are rendered. Enables management to track revenue generated for each type of product and service.
Service_Description	Dimension	Describes the attributes of each dial-up or dedicated access service. Enables management to track revenue and subscribers for each type of service over time.
Subscriber	Dimension	Identifies and describes the person or entity who subscribes to the services offered by the CSP.
Subscriber_Services	Subset	Allows for subscribers that have more than one service.

**FIG. 21A**

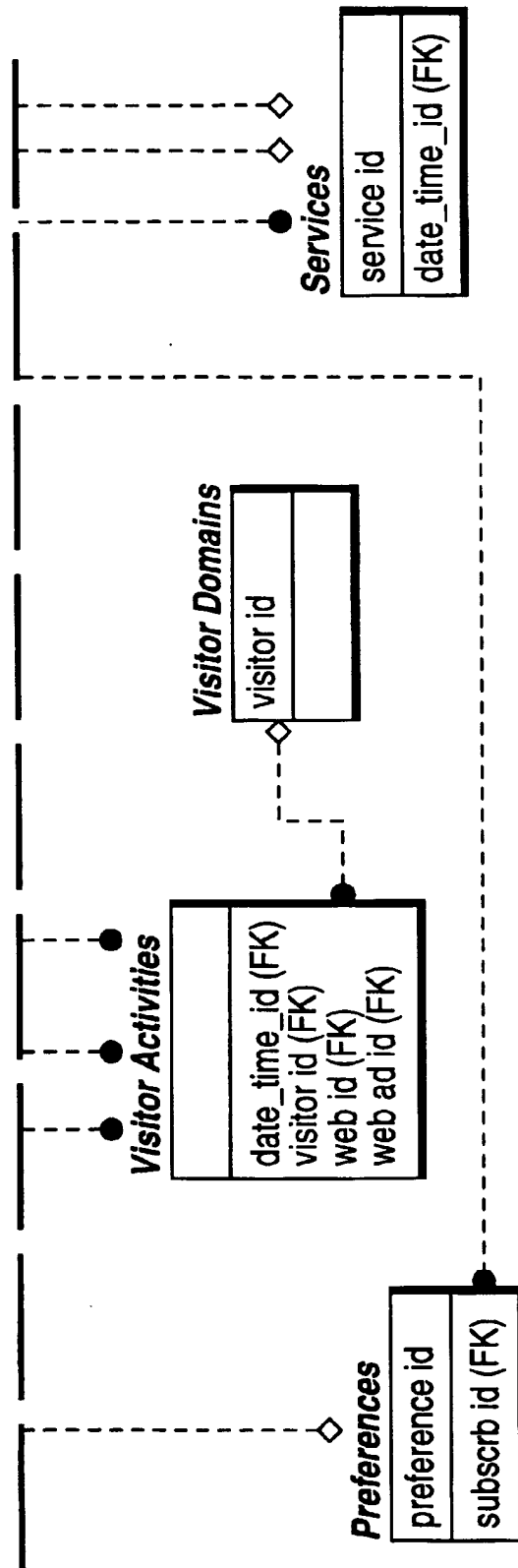


FIG. 21B

# COLLECTION AND INTEGRATION OF INTERNET AND ELECTRONIC COMMERCE DATA IN A DATABASE DURING WEB BROWSING

## BACKGROUND OF THE INVENTION

### 1. Field of the Invention

The present invention is generally related to a computer architecture and method for integrating data elements from Internet service providers (ISP) and commerce service providers (CSP) operational database(s) into predetermined format for supporting collection of Internet and/or electronic commerce data, and more particularly, to a computer architecture and method for integrating data elements from commerce service providers's operational database(s) into predetermined format for supporting collection of Internet and/or electronic commerce data over or from the World Wide Web for ISPs and CSPs.

### 2. Background of the Related Art

More and more people are using the Internet as a method of communicating, advertising and shopping for and purchasing goods. The sale of Internet services is growing at an amazing rate. The number of projected users into the year 2000 is dramatically impacting the communications industry both from the standpoint of an opportunity to realize new business and as a concern due to the potential loss of traditional revenue sources. FIG. 1 illustrates this dramatic increase sales in World Wide Web (WWW or web) servers, a summary of which is presented below:

WORLD WIDE WEB SERVER SALES FORECAST					
	1995	1996	1997	1998	1999
Intranet	475	2,673	5,483	9,210	13,133
Internet	621	979	1,410	1,777	2,159
Total	1,096	3,652	6,893	10,987	15,292

The explosive growth in PCs, Servers and Internet related software has cultivated a need for companies to better understand their customer's needs. To better understand these needs, many gigabytes of data must be collected and analyzed to arrive at the best way to service the customer.

Market and industry analysts alike, believe that the Internet will prove to be the most significant innovation in modern history since the light bulb and automobile. The method in which we perform daily business operations will be changed forever due to this new technology. Many technology based companies in the computer industry are scrambling to outline new products and services using and exploiting the Internet as a vehicle to increase market share and revenue, while increasing productivity and cutting operational costs. FIG. 2 is an illustration of the estimated growth in web users over the next several years.

In an effort to meet the above needs of digesting the vast amounts of information on the web, companies have designed many browsers and millions of web pages to access, retrieve and utilize this information. In addition to the Internet, companies have set up local "intranets" for storing and accessing data for running their organizations. However, the sheer amount of available information is posing increasingly more difficult challenges to conventional approaches.

A major difficulty to overcome is that information contained on the web or web pages are often dispersed across

the network at many sites. It is often time-consuming for a user to visit all these sites. One conventional approach used to access this information more effectively is called a search engine. A search engine is actually a set of programs accessible at a network site within a network, for example a local area network (LAN) at a company or the Internet and World Wide Web. One program, called a "robot" or "spider," pre-traverses a network in search of documents and builds large index files of keywords found in the documents.

A user of the search engine formulates a query comprising one or more keywords and submits the query to another program of the search engine. In response, the search engine inspects its own index files and displays a list of documents that match the search query, typically as hyperlinks. When a user activates one of the hyperlinks to see the information contained in the document, the user exits the site of the search engine and terminates the search process.

Search engines, however, have their drawbacks. For example, a search engine is oriented to discovering textual information only. In particular, they are not well-suited to indexing information contained in structured databases, e.g. relational databases, voice related information, audio related information, and the like. Moreover, mixing data from incompatible data sources is difficult in conventional search engines.

Another disadvantage with conventional search engines is that irrelevant information is aggregated with relevant information. For example, it is not uncommon for a search engine on the web to locate hundreds of thousands of documents in response to a single query. Many of those documents are found because they coincidentally include the same keyword in the search query. Sifting through search results in the thousands, however, is a daunting task.

Accordingly, we have determined that there is a need to be able to effectively collect the data, store the data and/or provide useful marketing information indicative of events occurring on the web. For example, data which indicates where a user has been in prior sessions may be useful in designing future products accessible via the web. We have also determined that there is a need for an architecture and method used to support and analyze Internet and/or electronic commerce data over or from the World Wide Web for CSPs and ISPs.

We have further determined that there is the need for an architecture and method used to correlate user, application, and access functions, and to store Internet and/or electronic commerce data in an efficient and multi-dimensional method. We have also determined that there is a need to provide a tool set that can easily communicate with, or become a subset of, an existing scaleable data warehouse to provide Internet and/or electronic commerce marketing decision support. Unfortunately, conventional architectures and/or techniques are unable to organize and present this information in an efficient manner.

## SUMMARY OF THE INVENTION

To overcome the above-detailed disadvantages and shortcomings of the prior art, it is a feature and advantage of the present invention to effectively collect the data, efficiently store the data and/or provide useful marketing information indicative of events occurring on the web. It is another feature and advantage of the present invention to collect data which indicates where a user has been in prior sessions, and which may be useful in designing future products accessible via and for the web.

It is another feature and advantage of the present invention to provide an architecture and method used to support



and analyze Internet and/or electronic commerce data over or from the World Wide Web for CSPs/ISPs.

It is another feature and advantage of the present invention to provide an architecture and method used to correlate user, application, and access functions and then to correlate these functions with business operational functions.

It is another feature and advantage of the present invention to provide a tool set that can easily communicate with, or become a subset of, an existing scaleable data warehouse to provide Internet and/or electronic commerce marketing decision support.

In the attainment of the same, the present invention provides an architecture and method of supporting, collecting, storing, and analyzing Internet and/or electronic commerce data over or from the World Wide Web for ISPs and CSPs. The present invention also includes one or more computer systems employing the method, a computer readable memory storing the medium, and a method for storing the synchronization process in memory and in a storage device.

The present invention combines improved methods of data collection, translation and storage with enhanced analysis techniques to exploit the vast amount of information available through emerging uses of the web, and other internet-based systems. The results can be used as part of a customer's decision support system to allow better (e.g., more statistically valid) business decisions to be made faster.

The present invention concentrates on opportunities within, for example, the ISP sector of the telecommunications industry, and provides ISPs the insight into their customer's services usage, demographics, and behavior patterns. Through this insight, ISPs should be able to support their decisions in such areas as customer retention, service pricing and profitability, and target marketing campaign using analytical methods based on customer and market data. Furthermore, ISPs can provide the knowledge and capabilities to the hosted business customers in order for those customers to gain similar insight into their users to make effective and customer-driven business decisions.

In addition, the present invention is complementary with the current Internet Services Delivery Platform program. One environment where the present invention operates is where the consumer or business customer is able to utilize the CSP's/ISP's web-based resources through a combination of access and backbone networks which, in turn, are connected to, for example, the global Internet to interface with the rest of the internet resources. The access networks consist of many Point of Presences (POPs) which provide dial-up (e.g., analog/digital, ISDN) access service or dedicated digital (e.g., Frame Relay, ISDN) access service using either modem pool technologies and/or access server/router for the customers.

The access networks also provide Dynamic Host Communication Protocol (DHCP) services, a Proxy Cache Server, and a firewall to facilitate the service administration, address translation, security authentication and/or authorization, and web operation efficiency. The backbone networks, consisting of switches and routers, optimize the transport between the access networks and the Internet Service Center. Internet Service Centers may provide Web Hosting Service and Electronic Commerce Applications.

The present invention provides an architecture supporting value-added services to improve the operation and services of ISPs/CSPs. The present invention leverages data warehouse and data mining techniques augmented with networking elements and web technologies. The goal of the present

invention is to provide the customers with dynamic, flexible, and adaptable interfaces and systems to fully utilize the interactivity and the tracking of the interactivity of the new Web medium. As a result, decision makers will be able to access decision support information through a secure network in a simple (ease of use), timely (fast enough), usable (valuable) and personalized (customized) fashion. Furthermore, this information will allow the customers to make intelligence decisions and take timely actions to achieve their business goals.

The present invention also is a decision support solution targeted at an Internet Service Provider (ISP) and Commerce Service Provider (CSP). CSPs and ISPs are service providers who offers a full range of value-added services such as electronic commerce applications. Its purpose is to help ISPs/CSPs to better manage their business, customer base, and services by leveraging areas of expertise in data warehousing and data retrieval.

The present invention was designed in view of the following goals which we have determined:

1. Understand electronic commerce over the Internet.
2. Understand the dynamics of the ISP/CSP business.
3. Determine the type of decision support needed to effectively run an Internet-based service business.
4. Understand how to leverage data warehouse and data storage and retrieval assets into this environment.

The present invention is a data warehouse—data storage and retrieval solution that is tailored to internet-based applications and services. One of its primary purposes is to allow ISP/CSP decision-makers to exploit the intelligence contained in datasets derived from the following environments: ISP/CSP business operations.

Internet or Web-based applications.

Web-based commerce applications.

Additionally, the decision-support information derived for specific applications can be provided by ISPs/CSPs to their electronic commerce clients as a service. The "discovery" of knowledge from each dataset is beneficial to ISPs/CSPs and their customers for the deep insights that it provides into business dynamics and customer behavior.

The scope of WebMining includes a data warehouse that extracts data from web logs as well as various business operational data bases. The warehouse, in turn, sources three datamarts: one focused on business operations, one on the CSPs customers and one on electronic commerce applications and on-line content. On-line content, in this context means, applications provided by the CSP as a service to his/her subscribers. The integration of web data to business data is key.

WebWarehouse and datamarts provide information in a business context, not a web site context. This is an important distinction. Although the system logs from one source of input to the warehouse, the data extracted from them is refined and translated into a business context before it is correlated with other operational data (customer, billing, product, etc.) and put into the warehouse. In other words, the warehoused data is refined to represent products being sold through electronic commerce applications, advertisements being seen by customers, specific content being viewed by customers, etc., not URLs. The extraction utilities and translation libraries that provide content for the warehouse will be as important to our offering as the decision support functions at the back end.

The systems elements of the present invention include:

1. mapping module: systems to collect, interpret, analyze, translate, refine and correlate customer Internet usage,

Web/Internet application data, electronic commerce application data and other operational data.

2. storage module: Scaleable Data Warehouse to structure, organize, store and manage the data.

3. data visualization: business decision-support solutions based on knowledge analysis and discovery of the data contained in the storage module.

The benefits of the present invention include:

1. providing decision support and adding operational information for the ISP/CSP to enable new/improved value-added services, allow the implementation of equitable and value-based pricing, achieve better quality of service, manage capacity, and add bill-back capabilities for charge back scenarios.

2. providing to the ISP's/CSP's customers valuable information such as decision support for proactive targeted marketing, usage and customer preferences feedback on the web applications, usage information for business access, direct decision support access for specific ISP/CSP customers' applications.

The WebWarehouse designer of the present invention provides a core design that is tailored for the decision-support functions. Extensions and/or changes to this design, however, may be optionally required to accommodate the ISP's/CSP's operational data and, perhaps, any additional decision-support requirements that surfaced during the data discovery phase. WebWarehouse designer contains two main components: a logical data model and a formatted library. The logical data model provides overall descriptions of the WebWarehouse. This description determines how well the WebWarehouse can integrate additional datasets and how effective the access of WebWarehouse will be for queries and reports. The Formatted File Library provides the intermediate classifications such as process characterizations, customer descriptions, preference determination, and behavior patterns. These classifications can beneficially be reused in different technical process and different customer problems.

In accordance with one embodiment of the invention, a system integrates data elements from an operational database of an Internet Service Provider (ISP) and Commerce Service Provider (CSP) into a predetermined format for supporting collection of the Internal and/or electronic commerce data. The system includes a database storing the Internet and/or electronic commerce data for interrogation by the ISP/CSP, and a user station including a design database. The design database includes a logical data model providing a description of the database facilitating the integration of the Internet and/or electronic commerce and operational data and facilitating query and report access of the database. The design database also includes a formatted file library providing classifications including at least one of process characterizations, customer descriptions, preference determinations, and behavior patterns, the classifications being reusable for at least one of different technical processes and different customer problems. As optional, a user interface is also provided which is capable of providing the functions of query access and report access of the design database.

One or more computer memory products are also provided. One or more computers memory stores, and one or more computers executes a program including the above process steps.

There has thus been outlined, rather broadly, the more important features of the invention in order that the detailed description thereof that follows may be better understood, and in order that the present contribution to the art may be

better appreciated. There are, of course, additional features of the invention that will be described hereinafter and which will form the subject matter of the claims appended hereto.

In this respect, before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of construction and to the arrangements of the components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced and carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein are for the purpose of description and should not be regarded as limiting.

As such, those skilled in the art will appreciate that the conception, upon which this disclosure is based, may readily be utilized as a basis for the designing of other structures, methods and systems for carrying out the several purposes of the present invention. It is important, therefore, that the claims be regarded as including such equivalent constructions insofar as they do not depart from the spirit and scope of the present invention.

Further, the purpose of the foregoing abstract is to enable the U.S. Patent and Trademark Office and the public generally, and especially the scientists, engineers and practitioners in the art who are not familiar with patent or legal terms or phraseology, to determine quickly from a cursory inspection the nature and essence of the technical disclosure of the application. The abstract is neither intended to define the invention of the application, which is measured by the claims, nor is it intended to be limiting as to the scope of the invention in any way.

The above objects of the invention, together with other apparent objects of the invention, along with the various features of novelty which characterize the invention, are pointed out with particularity in the claims annexed to and forming a part of this disclosure. For a better understanding of the invention, its operating advantages and the specific objects attained by its uses, reference should be had to the accompanying drawings and descriptive matter in which there is illustrated preferred embodiments of the invention.

#### BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention and the advantages thereof, reference is now made to the following descriptions taken in conjunction with the accompanying drawings in which like numbers designate like parts, and in which:

FIG. 1 is an illustration of the dramatic increase sales in World Wide Web servers;

FIG. 2 is an illustration of the estimated growth in web users;

FIG. 3 is an illustration of an Internet Services Delivery Platform program indicative of one environment where the present invention operates;

FIG. 4 is a high-level block diagram of a computer system with which the present invention can be implemented;

FIG. 5 is a diagram of a network in which the present invention can be implemented;

FIG. 6 is a flow chart of the framework of the decision analysis process of the present invention;

FIG. 7 provides a functional view of the system's elements in an Internet environment;

FIG. 8 provides an expanded view of the architecture of the present invention;

FIG. 9 is an illustration of the various data sources and outputs for the present invention;

FIG. 10 is a detailed illustration of how the components interact with each other and the customer environment in the present invention;

FIG. 11 provides an overview of the architectural components in various configurations of the present invention;

FIG. 12 is a table providing additional hardware, software, database, and information tool options that may be used in conjunction with the present invention;

FIG. 13 is a detailed illustration of the data discoverer architecture in accordance with the present invention;

FIG. 14 is an illustration of a functional architecture of the data discoverer in accordance with the present invention;

FIGS. 15–15A are illustrations of the architecture of the WebWarehouse designer of the present invention;

FIG. 16 is an illustration of the WebWarehouse data sources for the services portfolio management, subscriber base management, and application management;

FIG. 17 is a table key indicating relationships and table types;

FIG. 18 is a table key indicating different types of tables;

FIG. 19 is an exemplary data structure indicating subscriber service metrics data stored by the WebWarehouse designer and received from the data discoverer process;

FIG. 20 is an exemplary table indicating subscriber service metrics data stored by the WebWarehouse designer and received from the data discoverer process; and

FIG. 21 is an exemplary table indicating the various environmental and business data administered and/or collected by the data discoverer process and placed in the data warehouse.

#### Notations and Nomenclature

The detailed descriptions which follow may be presented in terms of program procedures executed on a computer or network of computers. These procedural descriptions and representations are the means used by those skilled in the art to most effectively convey the substance of their work to others skilled in the art.

A procedure is here, and generally, conceived to be a self-consistent sequence of steps leading to a desired result. These steps are those requiring physical manipulations of physical quantities. Usually, though not necessarily, these quantities take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared and otherwise manipulated. It proves convenient at times, principally for reasons of common usage, to refer to these signals as bits, values, elements, symbols, characters, terms, numbers, or the like. It should be noted, however, that all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities.

Further, the manipulations performed are often referred to in terms, such as adding or comparing, which are commonly associated with mental operations performed by a human operator. No such capability of a human operator is necessary, or desirable in most cases, in any of the operations described herein which form part of the present invention; the operations are machine operations. Useful machines for performing the operation of the present invention include general purpose digital computers or similar devices.

The present invention also relates to apparatus for performing these operations. This apparatus may be specially constructed for the required purpose or it may comprise a

general purpose computer as selectively activated or reconfigured by a computer program stored in the computer. The procedures presented herein are not inherently related to a particular computer or other apparatus. Various general purpose machines may be used with programs written in accordance with the teachings herein, or it may prove more convenient to construct more specialized apparatus to perform the required method steps. The required structure for a variety of these machines will appear from the description given.

#### BEST MODE FOR CARRYING OUT THE INVENTION

The present invention provides an architecture and method of supporting, collecting, storing and analyzing Internet and/or electronic commerce/transactions over or from (hereinafter “over” or “from” are interchangeable) the World Wide Web for ISPs and/or CSPs. The present invention also includes a computer system employing the method, a computer readable memory storing the medium, and a method for storing the synchronization process and in a storage device. The present invention provides tools that provide the necessary collection and analytical capabilities to derive knowledge from raw data collected from the web. In addition, the present invention provides scalability of processing and storage capacity, to meet a customer’s growing client needs.

In the following description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be apparent, however, that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are shown in block diagram form in order to avoid unnecessarily obscuring the present invention.

The present invention combines improved methods of data collection and storage with enhanced analysis techniques to exploit the vast amount of information available through emerging uses of the web, and other internet-based systems. The results can be used as part of a customer’s decision support system to allow better (e.g., more statistically valid) business decisions to be made faster.

The present invention concentrates on opportunities within, for example, the ISP sector of the telecommunications industry, and provides ISPs the insight into their customer’s services usage, demographics, and behavior patterns. Through this insight, ISPs should be able to support their decisions in such areas as customer retention, service pricing and profitability, and target marketing campaign using analytical methods based on customer and market data. Furthermore, ISPs can provide the knowledge and capabilities to the hosted business customers in order for those customers to gain similar insight into their users to make effective and customer-driven business decisions.

The present invention provides a complete data warehouse/decision support solution for Internet and/or electronic commerce-based applications and services. Its features include:

A modular design that scales in size as well as functionality.

It runs on a series of platforms ranging from single processors to massively parallel systems.

A full range of design and implementation tools.

Extraction and transformation programs tailored for the Internet and/or electronic commerce operating environments.

A choice of relational database management systems that includes MS SQL Server, Oracle, and Teradata.

A comprehensive selection of decision-support capabilities.

The present invention's modular design, hardware platforms and tools enable customers to select a decision-support system that is consistent with the financial maturity of their business. As the business grows and decision-support needs become more complex, the platforms described herein can be upgraded in size as well as functionality.

The present invention provides decision-support capabilities to help ISP's/CSP's manage their business from several perspectives:

Providing high quality Internet Service to subscribers as a foundation for higher value, more profitable services.

Managing the customer base to optimize profits and value.

Providing on-line content and services as a value-added incentive to attract and retain communities of subscribers, provide sources of advertising revenue, and potentially, to attract commerce clients.

Providing decision-support information, as a potential fee-based service, to business clients to help them manage their Internet, web or electronic commerce applications that are being hosted by the ISP/CSP.

Since a specific ISP/CSP may not be involved in every aspect of the business described above, the present invention's decision-support capability is modularized to allow a ISP/CSP to select the functionality that make sense for his/her business.

The present invention's modular structure allows Internet and/or electronic commerce-based applications and services to be managed from three perspectives:

1. Business Operations Management to monitor and understand:

Key business metrics (revenue, expenses, profitability, subscriber base).

Factors that impact quality of service.

The impact of marketing initiatives or competitive influences.

2. Customer-Base Management to understand the relative value, behavioral characteristics, and preferences of individual subscribers or customers.

3. Electronic Commerce Management to measure the success of:

Electronic-Commerce Applications—applications provided by the CSP's/ISP's or their business clients to conduct commerce over the internet.

On-line content—applications provided by the CSP/ISP to add value to the service offering.

In addition, the present invention is complementary with the current Internet Services Delivery Platform program. One environment where the present invention operates is illustrated in FIG. 3. Within this environment, the consumer or business customer 4, 6, 8 is able to utilize the CSP's/ISP's web-based resources through a combination of different access and backbone networks such as ISDN digital access, analog dial-up access, and/or dedicated access lines described below. These networks are, in turn, are connected to the Global Internet 2 via, for example, an internet service center 10 to interface with the rest of the internet resources.

The access networks consist of many Point of Presences (POPs) 12, 14, 16 which provide dial-up (analog/digital, ISDN) access service or dedicated digital (Frame Relay, ISDN) access service using either modem pool technologies

and/or access server/router for the customers. The access networks also provide Dynamic Host Communication Protocol (DHCP) services, a Proxy Cache Server, and a firewall to facilitate the service administration, address translation, security authentication and/or authorization, and web operation efficiency. The backbone networks, consisting of switches and routers, optimize the transport between the access networks and the Internet Service Center 10.

Internet Service Centers 10 may also provide Web Hosting Service and Electronic Commerce Applications. One implementation of Internet Service Centers for ISPs/CSPs is a single site with multiple servers 18 (Web Server Cluster) interconnected through a campus network 20. In addition, Internet Service Centers optionally provide Authentication Service, Domain Name Service, Mail Service, News Service, Provisioning Service, and Network Management Service.

The present invention provides value-added services to improve the operation and services of ISPs/CSPs. The present invention leverages data warehouse and decision support techniques augmented with networking elements and web technologies. The goal of the present invention is to provide the customers with dynamic, flexible, and adaptable interfaces and systems to fully utilize the interactivity of the new Web medium. As a result, decision makers will be able to access decision support information from a secure network in a simple (ease of use), timely (fast enough), usable (valuable) and personalized (customized) fashion. Furthermore, this information will allow the customers to make intelligence decisions and take timely actions to achieve their business goals.

The benefits of the present invention include:

1. providing decision support and operational information for the ISP/CSP to enable new/improved value-added services, allow the implementation of equitable and value-based pricing, achieve better quality of service, manage capacity, and add bill-back capabilities for charge back scenarios.

2. providing to the ISP's/CSP's, and their customers' valuable information such as decision support for proactive, targeted marketing, usage and customer preferences feedback on the web applications, usage information for business access, direction decision support access for specific ISP/CSP customers' applications.

FIG. 4 is a high-level block diagram of a computer system with which the present invention can be implemented. FIG. 4 is a block diagram which illustrates a computer system 100 upon which an embodiment of the invention may be implemented. Computer system 100 includes a bus 102 or other communication mechanism for communicating information, and a processor 104 coupled with bus 102 for processing information. Computer system 100 also includes a main memory 106, such as a random access memory (RAM) or other dynamic storage device, coupled to bus 102 for storing information and instructions to be executed by processor 104.

Main memory 106 also may be used for storing temporary variables or other intermediate information during execution of instructions to be executed by processor 104. Computer system 100 further includes a read only memory (ROM) 108 or other static storage device coupled to bus 102 for storing static information and instructions for processor 104. A storage device 110, such as a magnetic disk or optical disk, is provided and coupled to bus 102 for storing information and instructions.

Computer system 100 may be coupled via bus 102 to a display 112, such as a cathode ray tube (CRT), for displaying

information to a computer user. An input device 114, including alphanumeric and other keys, is coupled to bus 102 for communicating information and command selections to processor 104. Another type of user input device is cursor control 116, such as a mouse, a trackball, or cursor direction keys for communicating direction information and command selections to processor 104 and for controlling cursor movement on display 112. This input device typically has two degrees of freedom in two axes, a first axis (e.g., x) and a second axis (e.g., y), which allows the device to specify positions in a plane.

The invention is related to the use of one or more computer systems 100 for supporting, collecting, and analyzing Internet and/or electronic commerce over and/or from the World Wide Web for ISPs/CSPs. According to one embodiment of the invention, computer system 100 performs the synchronization in response to processor 104 executing sequences of instructions contained in main memory 106. Such instructions may be read into main memory 106 from another computer-readable medium, such as storage device 110.

However, the computer-readable medium is not limited to devices such as storage device 110. For example, the computer-readable medium may include a floppy disk, a flexible disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, any other optical medium, punch cards, paper tape, any other physical medium with patterns of holes, a RAM, a PROM, an EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave embodied in an electrical, electromagnetic, infrared, or optical signal, or any other medium from which a computer can read.

Execution of the sequences of instructions contained in main memory 106 causes processor 104 to perform the process steps previously described. In alternative embodiments, hard-wired circuitry may be used in place of or in combination with software instructions to implement the invention. Thus, embodiments of the invention are not limited to any specific combination of hardware circuitry and software.

Computer system 100 also includes a communication interface 118 coupled to bus 102. Communication interface 108 provides a two-way data communication coupling to a network link 120 that is connected to a local network 122. For example, communication interface 118 may be an integrated services digital network (ISDN) card or a modem to provide a data communication connection to a corresponding type of telephone line. As another example, communication interface 118 may be a local area network (LAN) card to provide a data communication connection to a compatible LAN. Wireless links may also be implemented. In any such implementation, communication interface 118 sends and receives electrical, electromagnetic or optical signals which carry digital data streams representing various types of information.

Network link 120 typically provides data communication through one or more networks to other data devices. For example, network link 120 may provide a connection through local network 122 to a host computer 124 or to data equipment operated by an ISP or CSP 126. ISP or CSP 126 in turn provides data communication services through the world wide packet data communication network now commonly referred to as the "Internet" 128. Server 132, web server or proxy server, is used by ISP 126 for data collection purposes. Local network 122 and Internet 128 both use electrical, electromagnetic or optical signals which carry digital data streams. The signals through the various net-

works and the signals on network link 120 and through communication interface 118, which carry the digital data to and from computer system 100, are exemplary forms of carrier waves transporting the information.

Computer system 100 can send messages and receive data, including program code, through the network(s), network link 120 and communication interface 118. In the Internet example, a server 140 might transmit a requested code for an application program through Internet 128, ISP 126, local network 122 and communication interface 118. In accordance with the invention, one such application provides for the method as described herein. Of course, the present invention may also be implemented on multiple computer systems.

The received code may be executed by processor 104 as it is received, and/or stored in storage device 110, or other non-volatile storage for later execution. In this manner, computer system 100 may obtain application code in the form of a carrier wave.

FIG. 5 is a diagram of a network in which the present invention can be implemented. Referring to FIG. 5, depicted is a network 200 within which the present invention may be implemented. SmartEC server 250, according to one embodiment of the present invention, gathers information dynamically from one or more data sources, which may be located at different servers and may have incompatible formats, and outputs the information for the user according to predetermined criteria discussed below.

The predetermined criteria may be defined by human operators according to their own needs, purposes, and preferences as part of the configuration of the server. Alternatively, the predetermined criteria may be determined based on system-defined user privileges. For example, users with a higher clearance or security and/or may be able to obtain more information than users with a lower clearance or security. Multiple information models and visual representations may be defined for any server.

A user may access the server by executing a web browser at client 210. Web browsers are well-known in the art, and are readily available from such corporations as Netscape Communications Corp. and Microsoft Corp. In order to access the server, the user at client browser 210 activates a hyperlink having a URL (Uniform Resource Locator).

In the exemplary URL, the network address of the server is specified as "www.server.com" and the optional portion of the URL after the question mark (?) hold user specified parameters. When the hyperlink is activated, the web server 220 receives a request to initiate a session, specified by parameters embedded in the URL. In response, the web server 220 gathers information from one or more data sources.

The data or information sources can have incompatible formats, e.g. web page, relational database, spreadsheet, text file, etc. The data sources can be stored at a plurality of sites, for example, locally with respect to the web server 220, such as a hard disk at local storage 222, or externally at another site in the network, e.g. at mainframe 230. In fact, the data source can even be another, remote information web server 240.

The decision analysis process shown in FIG. 6 describes a framework of what is required to provide the business solution set. The architecture and process of the present invention includes the necessary data to solve specific business problems. As opposed to many traditional methods of collection and storage for data warehouse and mining, the present invention focuses on data collection and correlation of Web/internet information and data storage methods using Web technologies in distributed environments.

As a result, data collection and data storage processes S2, S4 are focused and continuous efforts, instead of general and discrete approaches. The present invention also allows the proper archive of the collected information for statistic and analytic manipulation. The present invention effectively and interactively presents the resulting business solutions.

Specifically, three aspects of the processes are provided by the present invention business solution set:

1. Find and collect: this process S6 must determine what data elements are required. It is also necessary to know whether the data is available in open, closed, or registered environments. In order to execute the capture and collection of the data, the ability to set policy and operation parameters must be provided to the user to allow options for data searching and collection based upon their dynamic and changing business needs. Furthermore, this process should also deal with when and how the data collection should be executed.

The whereabouts of the required data, sometimes, can be known or unknown, internal or external. The ability to locate the data and synthesize the multiple sources of data into useful information is important as the technology evolves and matures. The following data collection options are also provided:

real-time: the time between locating the data and collecting the data is negligible or about at the same time. These two activities are closely related and the actions are closely coordinated.

on-line: the time between locating the data and collecting the data is short because either (1) the data requires a certain duration to accumulate meaningful volume, or (2) the importance of the data collection is not imminent. These two activities are loosely related and actions are loosely coordinated.

off-line: the time between locating the data and collecting the data is long and these two activities are separate and can be unrelated events, and actions need not be coordinated at all.

2. Structure and store: this process S8 involves parsing, categorizing, cleansing, correlating, indexing, and formatting of the collected data. The processed data must be organized in a structured way (i.e., the data layout and segmentation). The intent of this process is to arrange the data and/or information in a form where useful long term and policy-related knowledge can be easily used and derived.

3. Analysis and discovery: this process S10 recognizes the patterns, trends, and exceptions of the data and/or the information based on statistic and analytic manipulation techniques such as clustering, artificial intelligence, etc. The goal is to provide the customers with recommended actions and/or the visual interpretation of the presented data for their decision making S12.

All these three processes are iterative in nature. The feedback process S14 among the processes, and also from the users and customers, will establish the foundation of the continuing improvement. In addition, all three processes require user-friendly interfaces to facilitate information exchange. These user interfaces are important parts of the present invention to allow solution navigation, iterative learning, and decision guidance.

The following outlines the system elements and preliminary data required for the present invention. The present invention intends to complement the Internet service delivery platform and provide the differentiation for our complete suite of Internet solutions for Internet Service Providers (ISPs) and/or Commerce Service Providers (CSPs) in the communication industry. The assumptions of the configuration are:

all systems are interconnected,

all systems (users, servers) are Web accessible.

all ISP specify data (servers, database and warehouse) are accessible on a secure network.

The approach we take in the present invention is to establish a generic solution framework, but with specific focus, at least initially, on the vertical industry (ISPs in the communication industry) and functionally on customer retention and profitability. The potential applicability for horizontal platform functionality requires further investigation.

Each of the processes (find and collect, structure and store, analysis and discovery) corresponds to a system component in this section. FIG. 7 provides a functional view of the system's elements in an Internet environment. The system's elements of the architecture of the present invention are:

1. tracking module (WebTrack) 300: systems to find, search, collect and capture detailed customer Web/Internet and/or electronic commerce usage and demographics data.

2. storage module (Web Warehouse) 302: Scaleable Data Warehouse to structure, organize, storage and manage detailed customer Web/Internet and/or electronic commerce data.

3. intelligent decision support tool (WebSmart) 304: business solutions based on knowledge analysis and discovery of detailed customer Web/Internet and/or electronic commerce data.

In addition, a general data warehouse 306 is also provided (or existing data warehouse can also be referenced) for storage of non-web specific data, as well as possibly some web specific data based on user preferences. Each of the above elements are responsively connected, directly or indirectly, to web server 308, providing access to the internet 310.

In addition, the presentation of the information is also important to the decision support process. It is a delivery media to distribute the information and intelligence to the decision maker in a timely fashion. Therefore, all these systems provide consistent, personalized, and easy to use customer interfaces. These user interfaces should also have the capabilities to operate in real-time, on-line, and/or off-line.

FIG. 8 provides an expanded view of the architecture of the present invention. In FIG. 8, the users define and formulate the business questions and determine the queries of the data collected. Three major decision support user categories of the present invention are:

1. ISP/CSP business managers: the present invention allows them to better understand their customers' needs and conduct effective marketing and sales support to achieve their business objectives.

2. ISP/CSP IT (Information Technology) Managers: the present invention enables them to obtain insight into the ISP's/CSP's IT operations and provide ways to improve the ISP's/CSP's services.

3. ISP/CSP business clients: For the ISP's hosted business clients, the present invention enables ISPs/CSPs to provide their business clients with detailed understanding about the clients' business needs.

These users should be able to use the present invention through any Web client or networked client with properly defined access privileges and procedures. The interface must allow dynamic interaction between the users and the systems to accommodate the new queries and questions resulting

from the changing business needs. This user input is handled according to the current Web and user interface technologies such as HTML forms and Common Gateway Interface (CGI) or a window GUI interface interacting with the other elements (WebSmart, WebWarehouse, SmartEAdmin) of the WebMining systems.

WebSmart 304 provides the mathematical Engine to perform the statistic and analytical manipulation of the data and obtain the knowledge and intelligence for business decisions by the SmartEC client 314.

WebWarehouse 302 is separated from the operational client-accessible systems such as Web servers and Web-Based databases, but is populated by data from these systems. It exists to make the data available for interrogation by the business users from the perspectives of an enterprise interest and associated business benefits. The data in the WebWarehouse is preferably time-stamped and associated with a defined period of time, and is usually subject-oriented such as customer, product, activity.

The change of this data in the WebWarehouse occurs as a result of decision support requirements, and update from one of the data sources 316. Therefore, the SmartecAdmin 312 must have the capabilities to determine and control the search request and information flow between the operational databases and the WebWarehouse. SmartEAdmin must also have the ability to initiate the structural change, addition, and deletion of data in the WebWarehouse in an appropriate time based on the decision support user's needs.

WebMap 300 is the vehicle to search and transform the information from the data sources 316, and deliver the information to the WebWarehouse for storage using the secure network. The data sources preferably includes the system logs (access logs, agent logs, mail logs, communication server logs, accounting logs and referer logs) on the Web servers, proxy server or mail server, and the user profiles (ISP/CSP customers, visitors, and transient travelers (pass-throughs)).

From the Web operational perspectives, three major sets of information for Web applications and Web access must be also considered to determine the collection of useful data: (1) who is visiting, (2) what they are looking at, and (3) what resources they are using. There are other data sources available related to DNS, DHCP, and the Network Management Systems (NMS) which facilitate Web operations. In addition, the data sources should also consider third party analysis applications (based on the Web data).

The data required for the WebWarehouse depends heavily on the scope and objectives of the business problems to be solved. A preliminary and generic example of the WebWarehouse intended for solving the problems an ISP/CSP may face is described as follows:

From the user perspective, there are, at least, two types of users: visitors and the ISP's/CSP's customers. The ISP's customers are those who subscribe to the ISP services by paying fees to the ISP or purchase something from the ISP/CSP. The visitors are those who access the services via other Web servers without payment to the particular ISP/CSP. The profile data related to the ISP/CSP customers should include, for example:

- company name
- company administrator name
- user name
- address including city, state, zip code
- phone number including day, night and fax
- user login name and/or IP address
- email address

chosen Point of Presences

credit card information including card type, number, name on the card

domain name

billing contact

user selection of ISP's services including dial-up service, business ISDN service, corporate email service, server hosting, corporate web services.

user chosen service plans and associated pricing information.

The visitor profile data, in addition to the domain name and IP addresses, depends on how much information the visited applications can entice the visitors to provide and what the visitors are willing to share with the ISP/CSP.

From the Web operational service perspective, three activities are correlated: user, application (including hosting and commerce services), and access. Web servers contain minimal information about access from the networking point of view. The capture of Web access activity requires interfaces to other network accessible systems such as modem pools and routers.

The user activity provides, for example, the following data;

domain names and IP addresses of the systems/users making the requests

login names and IP addresses of the client systems accessing the site

referrer lists from where users are accessing the server

most common visitor domains

most common external references

first time visitors

returning visitors

most frequent visitors

visitors per visitor category

visits per visitor category

user demographics including geography, age, income, professional.

where do people come from when accessing this page

where users or subscribers/visitors go from this page.

The activity provides, for example, the following data:

hits to the site

new contents for the site

number of files sent

hot hits (most frequent accessed files)

most popular pages

types of transactions

date and time of transactions

time spent on a particular page

average time spent on a page

how many leave before the first page arrives

details of most popular objects

details of most popular pages

least popular objects

least popular pages

most popular entrance pages

most popular content categories

next click analysis

visits per content category

visits to content combinations

In order to understand the efficiency of the ISP/CSP operation, it requires an underlying resource characteriza-

tion to inter-relate the above activities. There are three aspects of this characterization: what the structure is (necessary Web site application/service structure, and the relationships—one-to-one, many-to-one, many-to-many), what actually happens (logs), and how well the systems are performing (performance). This characterization (please note that this remains to be defined) is important in setting up a useful WebWarehouse and provide adaptable services for the present invention. Sample data related to resource utilization includes, for example:

- protocol used
- file sizes transferred
- response time
- peak traffic periods
- update interval
- most common browser types
- retrieval method
- error activities
- accounting information including billing records
- services used (for example, support).

The present invention can be standalone or co-exist with the SmartEC decisions server, data bases, and data warehouse. The present invention accommodates both NT and UNIX with appropriate systems tools and capacity to enable access to other web servers and internet-based systems or networked such as NMS and special application servers. The present invention operates, for example, in a pure ISP environment initially. The capabilities of the present invention are also extendable to the Mixed ISP and SmartEC (or ISP decision support) systems provider environment (such that SmartEC systems providers can offer professional services), the Intranet environment for business customers, and provide business customers using ISP/CSP hosting and Internet and/or electronic commerce services a logical perspective of how they are operating.

As depicted in FIG. 9, the data to populate WebWarehouse 302 originates from many sources. The value of the present invention lies in its ability to integrate business operational data with web-specific and Internet and/or electronic commerce data to provide decision support information about Internet/Web applications in a business, rather than a technical context.

WebMap or WebTrack 300 consists of the programs and database structures designed to:

- Collect data from the appropriate sources.
- Interpret, analyze, translate, and refine the source data in order to extract the required data sets to implement decision-support functions.

Correlate disparate data sets prior to populating WebWarehouse 302.

WebWarehouse 302 furnishes the information necessary to manage an Internet Service or an electronic commerce business from the perspectives outlined above. Different WebWarehouse views can be extracted to provide an information framework to drive specific decision-support efforts. WebWarehouse also includes tools that provide the appropriate logical design, physical design and documentation. Design templates are available to identify the required data elements from the customer's operational databases. There are also processes and tools to integrate customer-specific elements into WebWarehouse's core design. This architecture described herein may also be implemented as a datamart. A datamart is structured to support specific decision support needs, typically at a departmental level.

WebSmart 304 provides decision-support modules with feature sets that vary in the sophistication of the business

information being provided. In this context, there are several levels of decision-support provided by WebSmart.

functionality that provides a basic level of decision-support functionality

functionality that can be addressed by standard OLAP decision-support software. At this level, WebSmart provides a set of pre-formatted report templates, filters and metrics for each decision-support feature set. There are tool sets for each decision support software package supported by WebMining or SmartEC decisions.

functionality used to detect previously unknown, but potentially useful relationships between different types of data. Here, WebSmart provides a set of templates oriented to help solve specific business problems such as customer retention, market basket analysis, etc.

Additionally, Metadata is provided for each of the supported On-Line Analytical Processing (OLAP or Knowledge Discovery (KD) tools, as appropriate. Metadata defines the data views necessary to produce the outputs required for each decision support feature set. Each metadata package also includes templates to integrate customer-specific data into the standard design.

FIG. 10 is a detailed illustration of how the components interact with each other and the customer environment. The present invention works under three environmental factors. Customer engagement 320 provides key information on customer needs. These needs are influenced by how well customer operational systems 322 are running and how customer business actions 324 are executed. The customer engagement process has a major influence on the design. The components must interact with customer operational systems to determine the proper data model for WebWarehouse 302 and how the information can be collected and extracted for WebMap or WebTrack 300. Finally, the customer business actions depend heavily on what knowledge customers can get from the products and solutions.

WebSmart 304 is intended to provide the best possible knowledge for customers to conduct their business given a specific business problem. There are six key components in the present architecture:

- WebMining or SmartEC decisions Front-end 326 (Data Discoverer 328 and Offer/System Configurator 334 and Migration Planner 336).

- WebWarehouse Designer 330

- WebMap or WebTrack 300

- WebWarehouse 302

- WebSmart 304

- WebMining or SmartEC decisions Metadata Management Utilities 332.

The WebMining Front-end facilitates the interaction between WebMining solutions and customers. Data Discoverer automates the discovery process to enable service practitioners to systematically find the necessary information for defining customer problems via customer engagement. The output of the Data Discoverer will impact the WebWarehouse design and determine the effectiveness of the collection and extraction utilities of WebTrack.

The offer/systems configurator and migration planner function (WebMining Platform Selector) will help both customers and WebMining sales and service personnel to determine, depending on customer needs, what hardware, software and associated applications are required and how customers can move from one set of platforms to another based on the changing customer requirements.

WebWarehouse Designer contains two main components: logical data model 338 and formatted library 340. The



logical data model provides overall descriptions of what the WebWarehouse is. This description will determine how well the WebWarehouse can integrate additional datasets and how effective the access of WebWarehouse will be for queries and reports. Formatted File Library is a by-product of conducting standard Collection/Extraction Utilities 342 and Tracking/Mapping Utilities 344. The Formatted File Library provides the intermediate classifications such as process characterizations, customer descriptions, preference determination, and behavior patterns. These classifications can be reused in different technical process and different customer problems.

WebTrack collects and extracts information from customer operational systems. The collected and extracted information will be further manipulated to determine the identity and characterization of the users via Tracking or Mapping Utilities. Data Transformation Tools 346 load the data into the WebWarehouse based on the principles of the WebWarehouse Designer and the information discovered during the execution of Collection/Extraction Utilities and Tracking and Mapping Utilities.

WebWarehouse is the physical repository of SmartEC or WebMining information.

WebSmart provides the intelligence of the present invention. Decision Support Systems (DSS) tools 348 provide the access, query, and report capabilities using WebWarehouse. Mining tools 350 enable the present invention to discover hidden knowledge from existing data and information. Furthermore, Mining tools could establish systems model such that Prediction/Forecasting 352 capabilities can be realized. All this intelligence can be formulated such that the business impact can be clearly crystallized via Business Impact Report 354.

Since there are multiple systems and databases involved in providing WebMining solutions, it is important to properly manage, coordinate, and reconcile the different sets of metadata. Metadata Management Utilities serve as a control center of these different metadata for the present invention.

The above architecture is mainly examined from input and output process perspectives among physical components. The physical platforms required for the architecture described herein are described in FIG. 11. The present invention is capable of running on, for example, an NT or UNIX server and can be accessed by users through a browser or DSS/Data Mining Tools GUI interface via a network connection.

The system environment in FIG. 11 provides an overview of the architectural components in various configurations. DSS/Data Mining can run locally on the database platform or remotely on an NT or UNIX server attached to the database server via a network connection. FIG. 12 is a table providing additional hardware, software, database, and information tool options that may be used in conjunction with the present invention.

FIG. 13 is a detailed illustration of the Data Discoverer architecture in accordance with the present invention. The Data Discoverer 360 is a front-end tool used to automate as much as possible the discovery process on customer information and business rules. The Data Discoverer determines or protects the structure of the data stored in the WebWarehouse. Data Discoverer provides the following functions:

- deletes, adds, updates information stored in the WebWarehouse
- reports information to the customer to allow customer to review the information and provide feedback (e.g., which information customer wants to, or more interested in receiving, customized reporting)

provides a user interface for setting up new customers to collect information to begin webmining operation WebWarehouse 302 is populated with data extracted from two primary sources: the CSP's/ISP's business operational databases 364, 366 and various Internet or Web-based system files 368. The Internet/web data is relatively consistent from implementation to implementation. Minor variances will exist depending on whether the customer is running on UNIX or NT, the logging options selected by the CSP/ISP and, in some cases, by the access hardware. These variances will impact the utilities that load the warehouse, but not its overall structure. The business operational data, on the other hand, may be unique for each CSP/ISP. WebWarehouse design templates take this into consideration by providing a fixed design for the consistent data and indicating, specifically, the type of data that must be extracted from the CSP's/ISP's operational systems to provide the functionality provided by WebSmart. User interface 372 interfaces or interacts with process controller 374 to store the data in reference database 376. User interface receives the data from the external sources via a data entry module for input into the various areas of the reference database, including schema, decision support taxonomy, data collection controller, or reference library.

Each of these various database areas are organized to maintain the discovered data in accordance with a predetermined and expandable format. Accordingly, at the end of this phase, the structure and location of all source data required for the warehouse is defined via the reference database. During the data discovery phase, specific decision-support objectives are also established and calibrated against the mission statement of the business. The results of the data discovery phase are then used to determine what extensions are needed to the WebWarehouse and WebSmart core designs.

FIG. 14 is an illustration of a functional architecture of the data discoverer in accordance with the present invention. The goal of the data discoverer 360 is to collect operational and business data, and then store the collected data in a well defined structure. The completed structure, also known as metadata, will provide the necessary input to WebWarehouse Designer, described above.

In FIG. 14, the service practitioner 380 starts the data discovery with an interview questionnaire 382. The questionnaire provides a structure to collect:

- the data
- a description of the data
- the environmental information about the data (e.g., how long files are kept on server without deleting, how current is data refreshed, how information is updated, how data is manipulated over time period, what POP or physical location)
- volatility of the data.

The data received via the questionnaire is then parsed, manually or automatically, into environmental characterization/related data 386 described above, and business related data 384. Examples of business related data include, e.g., pre-paid user, how user purchases services and products, discounts, billing rates, free subscriptions, free areas/information on web page.

Information source requirements 388 are determined from, or considers, at least one of the environmental characterization 386 or business characterization 384 when determining or populating the appropriate core business rules 392 and the appropriate core data sources 394 that

fulfill the information source requirements. Information source requirements comprise, for example, identifies where, physical/logical location, the information, business/environment is located in a database.

Core business rules 392 comprise a minimum or predetermined set of rules used to administer the subscriber account, e.g., predefined rules for the customer to perform operations, discounts, rebates, use of subscriber information involving logon id per user, account id group of users, subscriber id, internal database id, and the like. Core data sources 394 comprise the minimum data set which each customer is required to supply to the user data, e.g., user logon id, valid account, and the like.

Decision support requirements 390 utilize the business characterization when parsing into incremental data sources 396 and incremental business rules 398. The decision support requirements customized support requirements, information needed by customer beyond the core set of report functionality, e.g., county of residence which is not normally collected. Core data sources 394 and incremental data sources 396 are then parsed into or pointed to the appropriate internet/web source related information 400 and operational sources 402.

Internet/web source related information 400 and operational sources 402 are subsequently stored in warehouse designer 330, along with one or more core business rules 392 and incremental business rules 398. In this phase, the data discoverer will obtain the core data set 394. The core data set includes, but is not limited to, files, schema, tables, report, data elements and relationships.

The results of the data discovery process is highly iterative and highly dependent on the customers' needs. The purpose of data discovery is to achieve a starting point for a systematic way of discovering customer information. Interview templates are provided to guide the user through the process of identifying and recording information pertinent to WebWarehouse that describes the customers' information systems.

#### Business Operational Systems

Database or file names

Location of database or file

Business rational for each database or file

Database table descriptions and relationships

Data element descriptions for database tables or files

Business rules that dictate database or file data integrity

Update policies and frequency

Data volumes

Data volatility

Internet system logs.

The data discoverer also collects the following type of information:

Locations of data collection points

Data volumes

Operating system type and version

File formats

... about the following server logs:

Radius log—records information about access server sessions

Access log—records information about requests to local web applications

Proxy log—records information about requests to remote web applications

Referrer log—records information about hyperlinks between applications. In context of an application, log indicates who is linking to it

#### Mail log

The Data Discoverer is a self-contained Windows-based application that is capable of being run on a laptop computer. Data discovery occurs on the customer's premises. Data collection templates and processes are provided to inventory or store:

The types of services provided by the CSP/ISP

Billing codes

Service codes

Subscriber identification codes

On-line content

Electronic commerce applications.

With these types of applications, each one has a mutually exclusive business operations environment, customer base and business rules.

Data collection templates and processes are provided to collect data pertinent to the customer's Internet service environment. Included are:

Location and attributes of POPs (Point of Presence)

Topology and capacity of access facilities

Topology and capacity of Internet facilities

Types, names and locations of servers.

Data collection templates and processes are provided for the function of:

Validating the appropriateness of WebMining provided decision-support functionality for the customer's specific business needs.

Identifying and describing any additional business needs and/or decision support functionality that will require extending WebMining's core design and decision support templates.

Information collected about the customer's operational environment, information systems, and business needs is formatted into readable data files that can be integrated into WebWarehouse metadata.

FIGS. 15-15A are illustrations of the architecture of the WebWarehouse designer of the present invention. WebWarehouse designer 330 provides a core design that is tailored for the decision support functions provided by WebSmart 304 or alternatives. Extensions and/or changes to this design, however, are required to accommodate the CSP's/ISP's operational data and, perhaps, any additional decision-support requirements and/or changes that surfaced during the data discovery phase. WebWarehouse designer 330 contains two main components: logical data model and formatted library. The logical data model provides overall descriptions of what the WebWarehouse is. This description determines how well the WebWarehouse can integrate additional dataset, how effective the access of WebWarehouse will be for query and report. Formatted File Library is a by-product of conducting collection/extraction utilities and tracking utilities. The Formatted File Library provides the intermediate classifications such as process characterizations, customer descriptions, preference determination, and behavior patterns. These classifications can be reused in different technical process and different customer problems.

Tables, relationships, keys and data elements are integrated in the WebWarehouse metadata database 406. Also integrated are the locations and descriptions of the source data described during data discovery, the business rules, the database integrity conventions and the decision support views. The result is metadata that describes, for example, the specific implementation of WebWarehouse that is administered by the WebWarehouse designer. Metadata is data about

data. Metadata keeps track of and defines what is in the data warehouse, as well as the sources of data feeding the warehouse.

WebWarehouse Designer is optionally a Graphic User Interface (GUI) based design tool that provides the appropriate logical, physical design and documentation for each data warehouse or datamart. Also provided are the design templates required to integrate the data elements from the CSP's/ISP's operational databases into the appropriate data warehouse or datamart. These tools are intended to be used by either services practitioners or the customer's information systems staff. The metadata is implemented, for example, in a standard RDBMS format that is compatible with the RDBMS used to implement WebWarehouse. Other data formats may also be used so long as a similar format is used or compatible between the WebWarehouse Designer and the WebWarehouse.

WebWarehouse Designer is capable of reading a dataset and integrating the data into the WebWarehouse metadata. The location and format of customer-provided data is recorded during the data discovery process. The Data Discoverer 360 tool produces an output dataset filtered via the delta dataset module 414 used to only transmit data that has been altered to the WebWarehouse Designer. That describes this information in a predetermined but expandable format.

In FIG. 15A, the Data Discoverer tool 360 is a user tool. The user can be customers or professional service practitioners. WebWarehouse Designer 330 is a database administrator tool.

In this case, the WebWarehouse Designer 330 contains a predetermined and/or customers' design extensions. The specific core design of metadata drive the contents of the Data Discoverer 360. In conjunction with the user (customer) interaction, Data Discoverer 360 produces the updated business requirements on Metadata. At this point, a manual and/or automated procedure can be chosen to allow the design extension of the core design or design changes to obtain a customized core design.

In FIG. 15, the Data Discoverer is integrated into the core design by the integration function 412. Administration/Management in user interface 404 provides for Database Administrator (DBA) control and status, as required, during the integration process. Provisions are provided to allow a database administrator or decision support specialist to interact with the system. Change Control function in the user interface 404 allows elements to be added to the metadata 406 manually by a database designer. This allows changes and enhancements to be made. Query Control in the user interface 404 provisions for queries and reports when the metadata is established. This will provide the WebWarehouse documentation and will provide a reference for the decision-support analyst when designing new queries, reports or models.

The translator 408 provides a standard function for access to metadata 406 by either WebSmart 304 or WebTrack 300. WebTrack 300 requires definitions of customer-provided and Internet/Web elements in order to manipulate datasets during the WebWarehouse load process. Query/report generator 410 interfaces with user interface to output reports in paper or electronic form. The metadata is extensible. The metadata contains a description of each warehouse element that is included in the WebWarehouse core design. The core design provides the information structure that supports WebSmart's pre-defined decision-support functionality, i.e., information source requirements described above. The core design provided with the system can be extended to implement any additional and/or changed functions desired by the

customer, i.e., decision support requirements described above in connection with the Data Discoverer. Also, data elements required from the customer's files or databases must also be identified and included.

Metadata will provide a comprehensive description of each WebWarehouse component and the sources of data used to populate Webwarehouse. The metadata comprises the following information:

Schema for Internet/Web data sources

Type of system logs

Description of each log

Location of each log

Data format for each log entry

Operational considerations (e.g., frequency of data collection).

Schema for operational data sources

Required data sources

Schema describes the table, files and/or data elements that must be added to WebWarehouse to enable the decision-support functionality provided by, for example, WebSmart

Table/file location

Table/file responsible party

Table/file description

Table relationships (if database)

Data element descriptions

Customer extensions (describes the data sources necessary to implement customer desired extensions to WebWarehouse functionality).

Table descriptions including Table type, Table name, Table description, Primary Key, Alternate key, Foreign key(s),

Data Element Descriptions including Field Name, Label, Description, Parent Table, Shared by Alias, Data type, Length, Decimal places, Input mask, Display format, Legal characters, Key type, Uniqueness, Required value, Null support, Edit rules, Comparison boundaries, Operations allowed, Values entered by, Default value, Range of values

Table Relationship Descriptions

Business Rules

Decision Support Views.

The core design of WebWarehouse includes data structures, tables and the core data elements required for each table. The material is organized by data warehouse or datamart. FIG. 16 is an illustration of the WebWarehouse data sources for the services portfolio management, subscriber base management, and application management. FIG. 19 is an exemplary data structure indicating subscriber service metrics data stored by the WebWarehouse designer and received from the data discoverer process. The data structure section illustrates the tables used in a given data warehouse or datamart, the connections between tables that are related to one another and the types of relationships between them. Also shown are the primary and foreign keys. A primary key uniquely identifies the record in which it exists. A foreign key in one table is an attribute whose values are the values of a primary key in another table. The foreign key establishes a link between the two tables.

FIG. 20 is an exemplary table indicating subscriber service metrics data stored by the WebWarehouse designer and received from the data discoverer process. The table section describes why each table is important to the functionality of the data warehouse or datamart. This core design reflects the

set of business requirements that is expected to be substantially common from customer to customer. This core design is extensible to accommodate differences in customers' operational environments and business needs. Because of this and the broad scope of any data warehouse data mining effort, it is expected that the designs shown in this section will be modified and refined throughout the customer trial process. FIG. 21 is an exemplary table indicating the various environmental and business data administered and/or collected by the Data Discoverer process.

Accordingly, as described above, the present invention combines improved methods of data capture with enhanced analysis techniques to exploit the vast amount of information available through emerging uses of the web, and other internet-based systems. The results can be used as part of a customer's decision support system to allow better (e.g., more statistically valid) business decisions to be made faster.

The many features and advantages of the invention are apparent from the detailed specification, and thus, it is intended by the appended claims to cover all such features and advantages of the invention which fall within the true spirit and scope of the invention. Further, since numerous modifications and variations will readily occur to those skilled in the art, it is not desired to limit the invention to the exact construction and operation illustrated and described, and accordingly, all suitable modifications and equivalents may be resorted to, falling within the scope of the invention.

What is claimed is:

1. In a computer system for supporting and analyzing Internet and electronic commerce data over the World Wide Web for service providers, the computer system including at least one user station transmitting a user request, at least one server, connectable to said at least one user station, said at least one server providing said at least one user station access to external data sources being in a plurality of different data formats, and at least one data collection and analysis system, a system of integrating data elements from an operational database of the service provider including at least one of an Internet service provider (ISP) and a commerce service provider (CSP) into a predetermined format for supporting collection and integration of the Internet and electronic commerce data, comprising:

a database storing the integrated electronic commerce and Internet data for interrogation by the service provider wherein the electronic commerce data is merchant transaction data recording a specific purchasing transaction and the Internet data is data collected during web browsing by a user; and

a user station, responsively connectable to said database and including a design database storing:

a logical data model providing a description of the database facilitating the integration of the plurality of different formats of Internet and electronic commerce data and facilitating query and report access of the database;

a formatted file library providing classifications including at least one of process characterizations, customer descriptions, preference determinations, and behavior patterns, the classifications being reusable for at least one of different technical processes and different customer problems; and

a user interface capable of providing the query access and report access of the design database.

2. A system of integrating data elements from an operational database of a service provider into a predetermined format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein the

electronic commerce data includes environmental data, and the environmental data includes at least one of how long files are kept on a server without deleting, how current is the data, how often is the data refreshed, how the data is updated, how the data is manipulated over time, the physical location, and the point of presence.

3. A system of integrating data elements from an operational database of a service provider into a predetermined and extendable format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein the Internet and electronic commerce data includes business data, and the business data includes at least one of pre-paid user services, how the user purchases services and products, user discounts, billing rates, free subscriptions, and free information.

4. A system of integrating data elements from an operational database of a service provider into a predetermined and extendable format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein the design database stores information source requirements, and the information source requirements include at least one of physical location and logical location where the information is located.

5. A system of integrating data elements from an operational database of a service provider into a predetermined format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein the design database stores core business rules, and the core business rules include a minimum predetermined set of rules used to conduct customer business operation including administering a customer account.

6. A system of integrating data elements from an operational database of a service provider into a predetermined format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein the design database stores core data sources, and the core data sources include a minimum predetermined data set which the customer is required to supply to perform the subsequent retrieval of the information.

7. A system of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein the Internet and electronic commerce data information is retrieved from a distributed on-line, and interactive environment including multiple client browsers.

8. A system of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 7, wherein the multiple clients or client browsers are capable of accessing a SmartEC decisions server storing information.

9. A system of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 8, wherein the multiple client browsers are capable of accessing a server storing information including customer support material and sales campaign schedules.

10. A system of integrating data elements from an operational database of a service provider into a predetermined but flexible format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein the logical data model and the formatted file library comprise metadata, and the metadata is used to WebTrack and define the Internet and electronic commerce data stored in said database and sources of the Internet and electronic commerce data feeding said database.

11. A system of integrating data elements from an operational database of a service provider into a predetermined but flexible format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein said system is capable of reading a dataset and integrating the dataset into said database as metadata.

12. A system of integrating data elements from an operational database of a service provider into a predetermined format for supporting collection of the Internet and electronic commerce data according to claim 1, further comprising a translator providing an access function to access the logical data model and the formatted file library stored in said design database.

13. A system of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein the logical data model and the formatted file library comprise metadata, and the metadata includes a description of elements stored in said database in accordance with a core design.

14. A system of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 13, wherein the core design provides an information structure that supports pre-defined decision support functionality.

15. A system of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein the logical data model and the formatted file library comprise metadata, and the metadata provides a comprehensive description of each component and source of the Internet and electronic commerce data used to populate said database.

16. A system of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 1, wherein the logical data model and the formatted file library comprise metadata, and the metadata comprises schema for Internet/web data sources, and schema for operational data sources.

17. A system of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 16, wherein the schema for Internet/web data sources includes type of system logs, description of each log, location of each log, data format for each log entry, and operational considerations.

18. A system of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the electronic commerce data according to claim 16, wherein the schema for operational data sources includes tables, files and/or data elements that must be added to said database to enable decision-support functionality.

19. A method of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection and integration of the Internet and electronic commerce data being in a plurality of different data formats using a computer, comprising the steps of:

- (a) storing the integrated electronic commerce and Internet data for interrogation by the service provider in a

database wherein the electronic commerce data is merchant transaction data recording a specific purchasing transaction and the Internet data is data collected during web browsing by a user;

- (b) storing in a design database a logical data model providing a description of the database facilitating the integration of the plurality of different data formats of the Internet and electronic commerce data and facilitating query and report access of at least one of the databases and the design database;

- (c) storing a formatted file library providing classifications including at least one of process characterizations, customer descriptions, preference determinations, and behavior patterns, the classifications being reusable for at least one of different technical processes and different customer problems; and

- (d) providing the query access and report access to the design database.

20. A method of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 19, wherein the logical data model and the formatted file library comprise metadata, and the metadata is used to track and define the Internet and electronic commerce data stored in said database and sources of the Internet and electronic commerce data feeding said database.

21. A method of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 19, wherein said method is capable of reading a dataset and integrating the dataset into said database as metadata.

22. A method of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 19, further comprising a translator providing an access function to access the logical data model and the formatted file library stored in said design database.

23. A method of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 19, wherein the logical data model and the formatted file library comprise metadata, and the metadata includes a description of elements stored in said database in accordance with a core design.

24. A method of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 23, wherein the core design provides an information structure that supports pre-defined decision-support functionality.

25. A method of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 19, wherein the logical data model and the formatted file library comprise metadata, and the metadata provides a comprehensive description of each component and source of the Internet and electronic commerce data used to populate said database.

26. A method of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim

19, wherein the logical data model and the formatted file library comprise metadata, and the metadata comprises schema for Internet/web data sources, and schema for operational data sources.

27. A method of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 26, wherein the schema for Internet/web data sources includes type of system logs, description of each log, location of each log, data format for each log entry, and operational considerations.

28. A method of integrating data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection of the Internet and electronic commerce data according to claim 26, wherein the schema for operational data sources includes tables, files and/or data elements that must be added to said database to enable decision-support functionality.

29. A computer program memory, storing computer instructions to integrate data elements from an operational database of a service provider into a predetermined and expandable format for supporting collection and integration of the Internet and electronic commerce data over one or

more networks, the integrated electronic commerce and Internet data being in a plurality of different data formats, the computer instructions including:

- (a) storing in a design database a logical data model providing a description of the database facilitating the integration of the plurality of different data formats of Internet and electronic commerce data and facilitating at least one of query and report access to the design database, the electronic commerce data is merchant transaction data recording a specific purchasing transaction and the Internet data is data collected during web browsing by a user;
- (b) storing a formatted file library providing classifications including at least one of process characterizations, customer descriptions, preference determinations, and behavior patterns, the classifications being reusable for at least one of different technical processes and different customer problems; and
- (c) providing the query access and report access to the design database.

\* \* \* \* \*

UNITED STATES PATENT AND TRADEMARK OFFICE  
CERTIFICATE OF CORRECTION

PATENT NO : 6,128,624

DATED : October 3, 2000

INVENTOR(S): CHIAN, LUO-JEN, GAO, P., PAPIERNIAK, K., and THAISZ, J.

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

Title page item  
[75] Inventors: add --Peter Gao, Woodcliff Lake, NJ --

Signed and Sealed this  
Twenty-fourth Day of April, 2001

Attest:

*Nicholas P. Godici*

NICHOLAS P. GODICI

Attesting Officer

Acting Director of the United States Patent and Trademark Office